JOURNAL
OF
APPLIED PHYSICS

JOURNAL OF APPLIED PHYSICS

Vol. 21, No. 9 SEPTEMBER, 1950

In This Issue

	Page
Thickness Measurement of Thin Films by Multiple Beam Interferometry G. D. Scott, T. A. McLauchlan, and R. S. Sennett	843
Distribution in Energy of Johnson Noise Pulses Ben R. Gossick	847
Note on the Use of Reiner's Equation for Recognition of Pseudoplastics from Their Flow Curves Vera Osman Krakauer	850
X-Ray Study of Thermally Induced Stresses in Microconstituents of Aluminum-Silicon Alloys John P. Nielsen and Walter R. Hibbard, Jr.	853
X-Ray Diffraction by Elastically Deformed Crystals J. E. White	855
Some Effects of Nozzle Design on the Diffraction of Electrons by Gases R. B. Harvey, F. A. Keidel, and S. H. Bauer	860
Studies with an Aerodynamically Instrumented Shock Tube A. Hertzberg and A. Kantrowitz	874
Magnetostriction of Permanent Magnet Alloys E. A. Nesbitt	879
Improved Ultra-Thin Sectioning of Tissue for Electron Microscopy James Hillier and Mark E. Gettner	889
Reflection and Transmission of Electromagnetic Waves by Thin Curved Shells Joseph B. Keller	896
Technique for Taking Motion Pictures of Electron Microscope Images Luther E. Preuss and John H. L. Watson	902
Motion Picture Studies of Electron Bombardment of Colloidal Crystals John H. L. Watson and Luther E. Preuss	904
Emission of Radiation from Nitric Oxide: Approximate Calculations L. E. Benitez and S. S. Penner	907
Generalized Microscopy and the Two-Wave-Length Microscope M. J. Buerger	909
Model of Hurricane Formation Herbert Riehl	917
Nuclei of Strain in the Semi-Infinite Solid Raymond D. Mindlin and David H. Cheng	926
Thermoelastic Stress in the Semi-Infinite Solid Raymond D. Mindlin and David H. Cheng	931
Generalized Response of Linear Systems for Arbitrary Initial Conditions John R. Moore	933
Letters to the Editor:	
Thermal Expansion Coefficient and the Melting Point of Cubic Elements M. E. Straumanis	936
Tracing and Interpretation of Asymmetrical Hysteresis Loops G. M. Ettinger	936
Mechano Chemistry of the Dispersion of Mercury in Liquids in an Ultrasonic Field Evelyn C. Marboe and W. A. Weyl	937
Barium Oxide on Tungsten Cathode Interface Eugene B. Hensley and John H. Affleck	938
Alternative Method for the Summation of Fourier Series I. F. Morrison	939
Creep-Time Law for Zinc Crystals E. P. T. Tyndall	939
Fracture Modes in High Purity Metals T. J. Agnor and M. E. Shank	939
Intensity Correction Factors for X-Ray Spectrometer Transmission Pole Figure Determinations Suzanne Van Dijke Beatty	940
Determination of Elastic Constants in Single Crystals with Especial Reference to Silver Chloride David L. Arenberg	941
Examination of Thin Overgrowths by Multiple Scattering of Electrons L. G. Schulz	942
Note on Thermoelectric Generation of Current Max B. Greeff	943
On the Negligible Changes Produced in the Stress-Strain Curve by Immersion of a Copper Single Crystal in Mercury J. B. Wachtman, Jr.	943
Index to Advertisers	xxvi

BOARD OF EDITORS
Elmer Hutchisson, *Editor*

Associate Editors
Howard H. Aiken
Otto Beeck
James Hillier
W. James Lyons
Philip Morrison
Eric Reissner
Raymond J. Seeger
Frederick Seitz
Charles S. Smith

Publication of the
American Institute of Physics

OFFICERS
George R. Harrison, *Chairman*
George B. Pegram, *Treasurer*

GOVERNING BOARD
J. W. Beams
C. Paul Boner
Wallace R. Brode
W. F. Fair, Jr.
J. B. Fisk
G. P. Harnwell
George R. Harrison
W. V. Houston
Rudolph Kingslake
Paul Kirkpatrick
Paul E. Klopsteg
F. W. Loomis
Philip M. Morse
J. R. Oppenheimer
George B. Pegram
Harold K. Schilling
Francis G. Slack
John C. Steinberg

Consultant on Printing
Melvin Loos

ADMINISTRATIVE STAFF
Henry A. Barton, *Director*
Wallace Waterfall, *Executive Secretary*
Theodore Vorburger, *Assistant Treasurer and Comptroller*
Ruth F. Bryans, *Publication Manager*
Betty H. Goodfriend, *Assistant Secretary*
Eileen B. Neuberger, *Circulation Manager*

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members and associate members of the American Institute of Physics	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $10.00. Single copies: $1.00 each.

Changes of address should be sent to the American Institute of Physics: **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 21, Number 9 September, 1950

The Thickness Measurement of Thin Films by Multiple Beam Interferometry*

G. D. Scott, T. A. McLauchlan, and R. S. Sennett
Department of Physics, University of Toronto, Toronto, Canada

(Received February 2, 1950)

A method is described for the measurement of the thickness of thin films using the multiple beam interference techniques. Only simple equipment is required, but reduction of the order of interference to as low as *one* makes possible the measurement of average thicknesses down to 10A. Applications have been made to evaporated films of six different metals and the method appears to be useful for all stable thin films in thicknesses up to several microns.

Possible errors in the method and the precautions which may be taken to avoid them are discussed.

INTRODUCTION

VARIOUS applications of multiple beam interferometry to the study of surface topography have been discussed by Tolansky.[1] One of the important applications is to the measurement of the thickness of thin films by determination of the height of the step made by the edge of the film when formed on an optically smooth substrate. The method involves covering the film and adjacent substrate with a highly reflecting coating and then measuring the resulting step in the coating. Donaldson and Khamsavi[2] have shown that for silver films with a silver coating this procedure gives the true thickness of the film. However, Avery[3] has noted that errors may be introduced in some cases, for example when chromium is used as the reflection coating for measuring a silver film.

Reflection fringes with white light as viewed in a spectroscope have been found to be the most satisfactory type of fringes for such measurements. Reflection rather than transmission fringes are used since then the reflection coating can be made sufficiently thick to eliminate a possible difference in the phase change upon reflection at the coating which is on the film compared to that which is on the substrate alone. Silver is the best metal for the coatings since it has the highest reflection coefficient, at least in the most convenient region of the spectrum. White light and a spectroscope are employed rather than monochromatic light since then optical flats are not required for the substrates and all measurements can be made simply from the wave-length drum of the spectroscope.

EXPERIMENTAL

The general arrangement of the equipment is shown in Fig. 1. The interference plates are mounted on the table of a small goniometer, the arms of which carry the source and the projector lens, L_2. Light from the source, which was a 6-volt, 32-cp incandescent lamp, after passing through the condensing lens L_1 is reflected normally to the interference plates by a half-silvered mirror. An image of the plates is formed on the slit of the spectroscope by the projector lens. The focusing of the condensing lens is not critical, but a convenient adjustment is to have an image of the source formed at the aperture of lens L_2. To ensure sharp fringes the effective angular aperture of the projector lens should be less than 1/15. The spectroscope was a Hilger constant deviation instrument with calibrated wave-length drum. A 12× eyepiece with single vertical cross hair was employed. The projector lens was a 32-mm microscope objective which produced an image of the interference plates enlarged 3× at the slit of the spectroscope. Thus the over-all lateral magnification was 36×.

The details of the interference plates are shown in Fig. 2. Standard 1 in.×3 in. microscope slides were used for these plates. The front slide is coated with silver to give a 1 percent to 2 percent transmission; the back slide is the substrate for the thin film to be measured. A step or channel is made in the film along the length of the

* Presented in part at the annual meeting of the E.M.S.A., at Washington (October, 1949).

[1] S. Tolansky, *Multiple Beam Interferometry* (Clarendon Press, Oxford, 1948).

[2] W. K. Donaldson and A. Khamsavi, Nature **159**, 228 (1947).

[3] D. G. Avery, Nature **163**, 916 (1949).

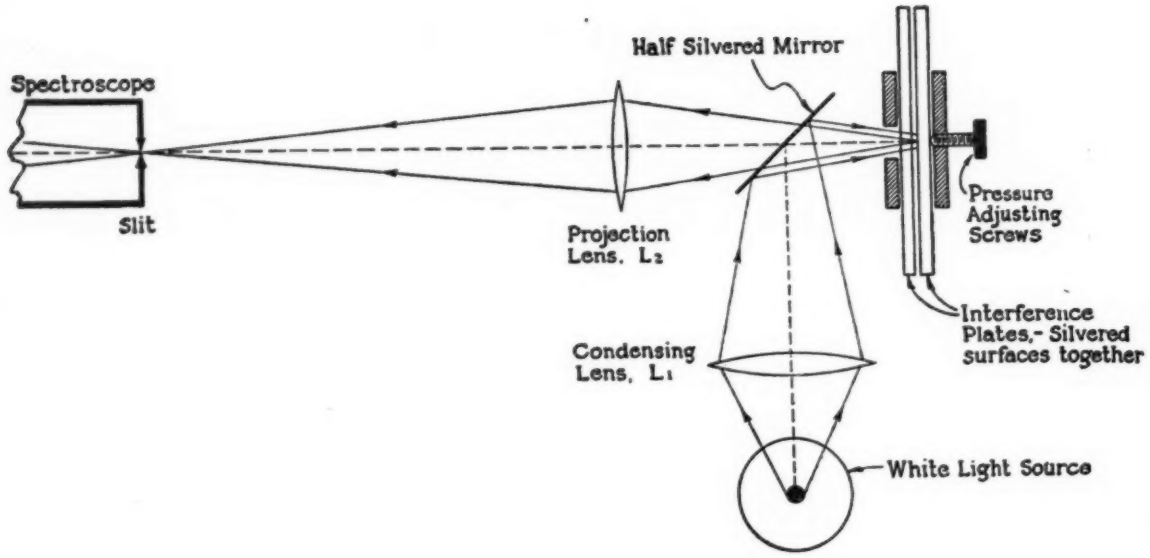


FIG. 1. Experimental arrangement for producing multiple beam interference fringes.

slide either during or after formation. The slide is then heavily coated with silver by thermal evaporation in vacuum. About 1000A is sufficient to eliminate the possible phase differences mentioned above. This silver surface thus has a step or channel, the depth of which corresponds to the thickness of the original film. It is important that the silver coatings have low optical absorption. In addition to the usual conditions for obtaining such coatings, it has been found that a rapid rate of deposition of the silver is also necessary.[4] The mounting which holds the plates on the goniometer table has a 3-mm aperture in the front and two small pressure-adjusting screws at the back. These screws control the separation of the plates.

The step or channel can be made in any convenient manner. It is desirable of course to have as sharp edges as possible. For metallic films, one method which was found quite satisfactory was to form a channel by gently drawing a not-too-sharp needle across the film. Some experience is required in order to remove all the metal yet not to scratch the glass substrate. The edges of the channel should be sharply defined and the bottom quite smooth. The slightest marring of the glass is readily detectable from the shape of the fringes. Another method is to form the film with the step or channel by using a suitable shield such as a narrow strip of metal foil. Both methods were found to give the same thickness measurements within the limits of error.

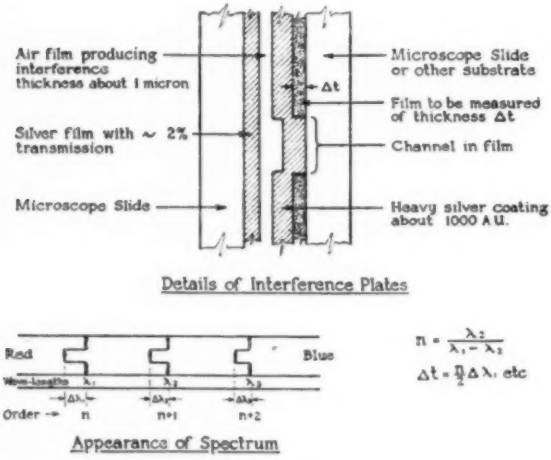


FIG. 2. Details of the coatings on the interference plates and a schematic diagram of the fringes from the plates.

A number of adjustments are required to obtain satisfactory fringes. The spectroscope is focused for monochromatic light and the slit width set for maximum resolution. The components as shown in Fig. 1 are aligned geometrically so that when the plates are in position a white light spectrum with interference fringes is observed in the spectroscope. Adjustments are made as follows.

1. Lens L_2 is focused to give sharp fringes.
2. The pressure screws are adjusted to make the fringes nearly vertical, at least in the region of the step.
3. Lens L_2 is moved laterally to locate the position where the fringes remain stationary. This corresponds to the area where the plates are exactly parallel and hence where the fringes have optimum sharpness and contrast.
4. The pressure screws are then set for the desired order and at the same time adjustment (2) is maintained, and (3) is checked.

While moving lens L_2 as in (3), fringes are viewed corresponding to various positions along the step. In this way local irregularities can be noted and avoided so that a final representative position is selected.

A schematic diagram of the spectrum observed is given in Fig. 2. The wave-length interval $\Delta\lambda$ corresponding to the height of the step Δl is measured directly, using the wave-length drum of the spectro-

[4] R. S. Sennett and G. D. Scott, J. Opt. Soc. Am. **40**, 203 (1950).

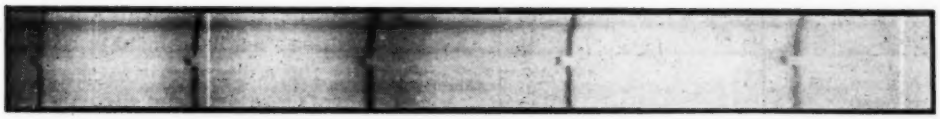


FIG. 3. Spectrogram of multiple beam interference fringes of orders 12 to 16. The mercury lines superimposed as wave-length references are the yellow doublet at 5770A and 5791A, the green line at 5461A, and the blue line at 4358A. The discontinuity in the fringe is caused by a channel in a silver film 115A thick.

scope. The order n for the fringe is determined as indicated by the formula in Fig. 2 and then Δt may be calculated: $\Delta t = (n/2)\Delta\lambda$. The spectrum of a typical fringe system at relatively high orders, *viz.*: 12 to 16, is shown in Fig. 3. Measurements were usually made at orders from 1 to 6. At order *one* for example, the "discontinuity" in the fringes of Fig. 3 appears as shown in Fig. 4. The mercury spectrum has been superimposed on these spectrograms merely to indicate the wave-length scale.

Since microscope slides are used in place of optical flats, the fringes are not in general straight; however, the use of the slides, which are relatively thin, makes it possible to obtain interference orders as low as *one* or *two* quite readily. Such low orders add greatly to the convenience with which readings can be made. It may be noted, for example, that for order *two* the wave-length interval at the step is just equal to the thickness of the film. If the fringe for order *two* is at the red end of the visible spectrum, the fringe of order *three* is in the blue, so that the order number can be immediately recognized. Furthermore it was observed that with microscope slides the fringes were about one-quarter the width of those obtained with ordinary optical flats. This presumably indicates that over distances of say 10 microns the microscope slides which are rolled or fire polished are smoother than the flats which are polished by an abrasive. Irregularities of 10 to 20A in the flats would account for the observed effect and it seems possible that these could be present in standard quality flats.

The fringes are produced by a very narrow area of the interference plates—the area corresponding to the image of the slit at the plates. With the arrangement described here, an individual setting on a fringe represents a position which averages heights over an area of about 5×50 microns on the plates. However, because of the necessary adjustments and the desirability of avoiding local irregularities, it is preferable to have available a film area of at least the order of 1 sq. mm.

It was found that within experimental error measurements of the depth of a step or channel were the same at all useful orders, i.e., from *one* to about *thirty*. For interference order *two* the wave-length settings on a fringe could be made in general to within 2A. Hence for a film 100A thick the measurement should be correct certainly within 5 percent.

A possible source of error in these measurements is the variation with wave-length in the phase change for reflection at a thick silver film. From a plot of $n\lambda$ against λ, it was found that this variation is quite small. The approximate correction for this effect in the measurement of the thickness Δt is given by: $\Delta t = (n/2)\Delta\lambda + [0.01(3)/2]\Delta\lambda$. It is evident that even at order *one* the correction is only $\frac{1}{2}$ percent and hence for the sharpness of fringes at present obtainable this error can be neglected.

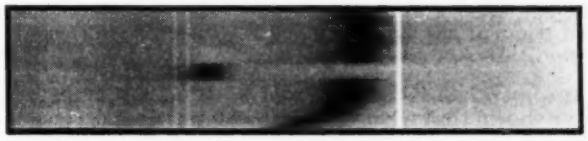


FIG. 4. The same channel as in Fig. 3, but at order *one*. The mercury yellow doublet and green line are superimposed on the spectrum.

As Tolansky[1] has emphasized, the method is capable of high resolution in the measurement of average heights. This fact is illustrated in Fig. 5 which shows a fringe at order *one* used to measure a silver film 15A in thickness.

DISCUSSION

The displacement of the discontinuity in the fringe certainly gives the depth of the channel or step in the silver coating, but whether this is precisely the same as the thickness of the original film must be considered in detail. It seems reasonable to assume that it is in the case of films which are continuous and of much the same nature as the glass substrate, such as Formvar films. However for films which are either aggregated or quite different in nature from the glass substrate, the assumption may not hold. Avery[3] has shown that when a chromium coating is used to measure silver films, the apparent thickness is about 30 percent greater than when silver is used as the coating. This may be explained by assuming that chromium does not fill the interstices of a silver film as completely as does silver itself. Alternately, it could also be assumed that the metal just adjacent to the glass is less aggregated for the chromium than for the silver. However, Donaldson and Khamsavi[2]

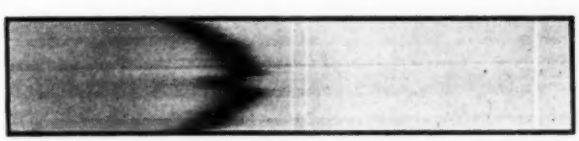


FIG. 5. The discontinuity in a fringe due to a silver film of average thickness 15A.

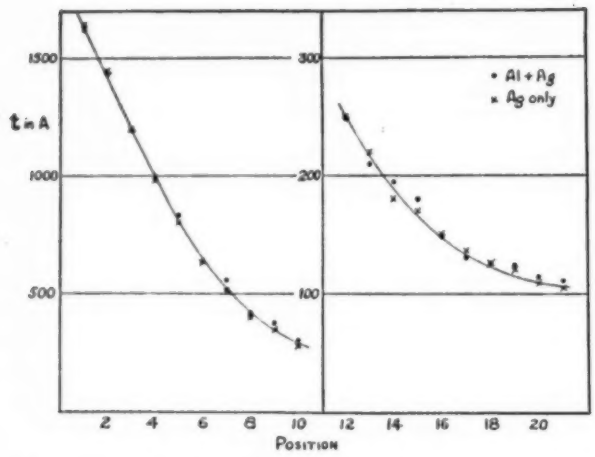


FIG. 6. Effect of an intermediate layer of aluminum on aluminum film thickness measurements using silver for the high reflection mirror.

have shown that when a silver coating is used for a silver film the measured thickness gives a density for the metal in the film equal to that of the bulk metal. This result is applied to films as thin as about 100A, which are known to be aggregated. Since electron diffraction shows that lattice spacings in thin films are essentially the same as in the corresponding bulk metals, it can be concluded that at least for the case of silver the multiple interference method yields the true average thickness for aggregated as well as continuous films.

For films other than silver the following method can be used in order to avoid errors due to the effects mentioned above. The film to be measured, A, is first covered with an intermediate layer, B, of the same material and deposited under the same conditions as A. The coating B is made sufficiently thick so that its top surface has the same structure over the substrate as over the film A. Finally the high reflection silver coating is added. It is evident that this procedure compensates for peculiarities (1) in the form of the material of the film in contact with the substrate and (2) the way in which the first layer of the silver coating may be deposited on the film material.

Thin films of the following metals, in addition to silver, have been measured in the range from 40A to 1500A: gold, copper, aluminium, antimony, and chromium. For each case at least two channels were made side by side through the metal films which varied in thickness along the length of the microscope slides. One channel was first coated with a thick film of the same metal, and then both were coated with a heavy silver film. For all the metals studied the thickness measurements were the same within the limits of error, either with or without the intermediate layer. As an example the results for aluminium are shown in Fig. 6. The measurements for the case of silver on chromium showed no discrepancy though Avery found that errors do occur when chromium is used as the coating for silver. It can be concluded that when silver is used as the reflection coating, the use of the intermediate layer is unnecessary for films of gold, copper, aluminium, antimony, and chromium.

It is essential to the method that the intermediate coating, or in the simpler case the silver coating, be of precisely the same thickness on the bare substrate as on the film. This condition is readily obtained when the films are deposited by thermal evaporation in vacuum, provided none of the incident atoms are reflected or scattered away from the surface which they strike. It is known that the distance of atomic migration over the substrate is much too small to be of importance in this connection. However, during the initial stages of the formation of these coatings, differences in the scattering of atoms from the films compared to the scattering from the substrates could occur. It is highly probable that this effect would only be important for films less than 50A in thickness. Because of the difficulties of making independent and accurate measurement of films of such a thickness, no experimental information on the importance of the effect is available. The possibility of such an error could be eliminated by forming the film on a substrate of the same material as the film itself.

CONCLUDING REMARKS

The method described here for the measurement of the thickness of thin films appears applicable to any stable thin film which can be formed on a smooth substrate and with a sharp edge. It is particularly suitable to film thicknesses in the range from 10A to 1000A, but can be used for films up to several microns in thickness. The precision at 100A is better than 5 percent.

The equipment required is simple and the results can be obtained quickly. After the film has been properly coated, about two minutes is usually sufficient time to make the adjustments and obtain a thickness measurement.

There are certain possible errors in the method, but in many cases these are not important. However each particular type of film and substrate should be considered independently to determine what precautions may be necessary.

Distribution in Energy of Johnson Noise Pulses

BEN R. GOSSICK*
NEPA Division, Fairchild Engine and Airplane Corporation, Oak Ridge, Tennessee
(Received August 22, 1949)

The distribution in energy of Johnson noise pulses (random thermal voltages in a resistor) is derived and checked against pulse-height measurements made with a linear amplifier and electronic counter. The following parameters are determined from the distribution function: (1) time of flight associated with a Lorentz mean free path, (2) the number of electrons which produce a pulse of a given height, and (3) the potential an electron falls through in a Lorentz mean free path.

DISCUSSION

NYQUIST'S equation[1]

$$d\langle V^2\rangle = 4kTRdf \tag{1}$$

applies to random thermal voltages across either an ohmic resistor or the radiation resistance of a Hertzian dipole. Bernamont,[2] and later Bakker and Heller,[3] using Lorentz electron theory and Fermi-Dirac statistics, obtained a proof of Eq. (1) in which the ohmic resistance appears explicitly. The distribution of random noise currents has been treated by Landon,[4,5] Norton,[5] Kac,[6] and Rice.[7] Norton draws upon Rayleigh's[8] solution for the amplitude distribution of a large number of unit vibrations with random phase.

The purpose of this paper is twofold: one, to present a solution for the Johnson[9] noise pulses by a method which is thought to be new, and two, to supply measurements of Johnson noise pulses made with a linear pulse amplifier and electronic counter. It is believed that such measurements have not been previously reported.

The distribution of noise pulses is analyzed here by Maxwell-Boltzmann statistics using a time-energy phase space. It is assumed that the resistor is in thermal equilibrium, and that the Johnson noise is a stationary random process. In considering the distribution of pulses by Maxwell-Boltzmann statistics, it is convenient to employ the familiar treatment of molecules as an analog. This analogy is given in Table I.

The probability of a distribution of pulses p_1, $p_2\cdots$ becomes

$$W = p!\prod_i (g_i)^{p_i}/p_i!. \tag{2}$$

As the number of pulses is large, Stirling's approximation is applicable, so that (2) can be rewritten as

$$W = (p)^p \prod_i (g_i/p_i)^{p_i}. \tag{3}$$

In order to obtain the most probable complexion, $\ln W$ is maximized by the method of Lagrangian multipliers. The following equations of condition are adopted:

$$\sum_i p_i - (t'/\tau') = 0, \quad \tau' \geqslant \tau, \tag{4}$$

$$\sum_i p_i - (t'/\tau) = 0, \quad \tau' < \tau. \tag{5}$$

$$\sum_i p_i E_i - E = 0, \tag{6}$$

where τ is the time of flight associated with the Lorentz mean free path, τ' is the response time of the electronic system, E_i is the energy of a pulse in the ith cell of phase space, and E is the total energy of all pulses occurring in the interval $0<t<t'$. Equations (4) and (5) state that the total number of pulses in the counting interval $0<t<t'$ is invariant, and this is an approximation requiring that $t'\gg\tau$, and $t'\gg\tau'$. Because of the limitations of electronics, Eq. (4) is normally applicable. Equation (5) is included for completeness to the theory, and because it seems possible that it could be used in special cases, e.g., for τ across a biased barrier of a semiconductor-metal rectifier. Equation (6) follows from the original assumption that the resistor is in thermal equilibrium. Solving for the number of pulses p_i in the ith cell of phase space yields:

$$p_i = g_i e^{-1-\alpha-\beta E_i}, \tag{7}$$

where β and α are the Lagrangian multipliers associated with (6) and (4), respectively. Following the molecular analog, (7) can be expressed as the differential:

$$dp^2 = Ae^{-\beta E}dEdt, \tag{8}$$

which gives the number of pulses with energy between E and $E+dE$ arriving between t and $t+dt$, where A is a constant to be determined directly.

TABLE I.

Molecules	Pulses
1. Container of volume V enclosing M molecules	1. Interval of duration t' enclosing p pulses
2. Dimensions of box are large compared to dimensions of molecules	2. Interval is large compared to pulse width
3. Statistical weight is proportional to $\Pi_i dx_i dp_i$	3. Statistical weight is proportional to $dEdt$

* Present address: Physics Department, Purdue University, Lafayette, Indiana.

[1] H. Nyquist, Phys. Rev. **32**, 110 (1928).
[2] H. J. Bernamont, Ann. de physique **71** (1937).
[3] C. J. Bakker and G. Heller, Physica **VI**, 262 (1939).
[4] V. D. Landon, Proc. I.R.E. **29**, 50 (1941).
[5] V. D. Landon and K. A. Norton, Proc. I.R.E. **30**, 425 (1942).
[6] M. Kac, Am. J. Math. **65**, 609 (1943).
[7] S. O. Rice, Bell Sys. Tech. J. **23**, 282 (1944); **24**, 46 (1945).
[8] Lord Rayleigh, *Theory of Sound* (Dover Publications, New York, 1945), Chapter II, 42a.
[9] J. B. Johnson, Phys. Rev. **32**, 97 (1928).

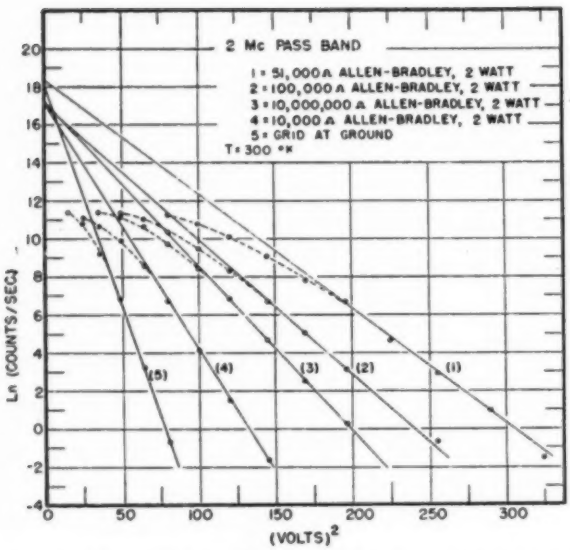


FIG. 1. (Volts)² refers to the square of the pulse-height selector setting at the A-1 amplifier. Hence, to relate these curves to (23) it is necessary to introduce the over-all gain G into (23) as follows:

$$\dot{p}_{V_0} = (1/\tau') \exp(-V_0^2G^2/2\langle V^2\rangle G^2)$$
$$= (1/\tau') \exp[-(\text{volts})^2/2\ (\text{mean square volts})]$$
$$\ln\dot{p}_{V_0} = \ln(\text{count/sec.})$$
$$= -(\text{volts})^2/[2\ (\text{mean square volts})] + \ln 1/\tau'.$$

The observable energy per pulse equals

$$E = V^2\tau'/2R, \tag{9}$$

where V is the amplitude of the voltage pulse. By means of (9) Eq. (8) can now be rewritten in terms of voltage as

$$d\dot{p}^2 = (A\tau' V/R)dVdt \exp(-\beta\tau' V^2/2R). \tag{10}$$

As the total number of pulses which arrive in the interval t' is the quotient t'/τ', it follows that

$$t'/\tau' = \frac{A\tau'}{R}\int_0^{t'}\int_0^{\infty} VdVdt \exp\left(\frac{-\beta\tau' V^2}{2R}\right). \tag{11}$$

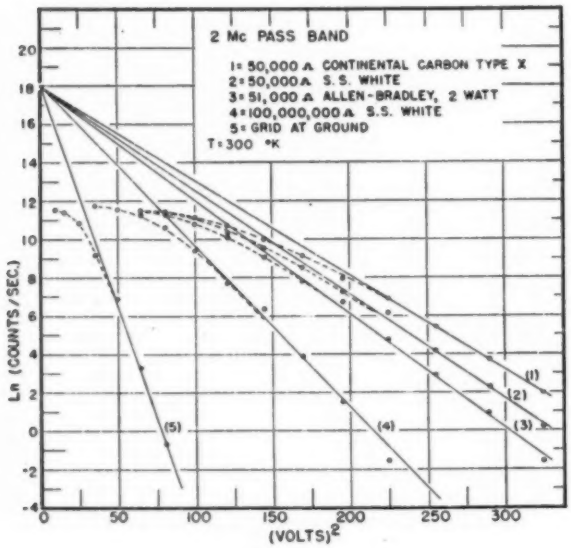


FIG. 2.

Solving (11) for A yields

$$A = \beta/\tau', \tag{12}$$

which can be used to rewrite (10) as

$$d\dot{p}^2 = (\beta V/R)dVdt \exp(-\beta\tau' V^2/2R). \tag{13}$$

As it was initially assumed that the Johnson noise is a stationary random process, the function of interest is the mean pulse rate, hereafter denoted by $\dot{p}$. Thus

$$d\dot{p} = (\beta/R)VdV \exp(-\beta\tau' V^2/2R) \tag{14}$$

gives the average rate of pulses with amplitude between V and $V+dV$.

The constant β is evaluated immediately in terms of the mean power $\bar{P}$.

$$\bar{P} = \frac{\beta}{R}\int_0^{\infty} EVdV \exp\left(\frac{-\beta\tau' V^2}{2R}\right), \tag{15}$$

$$= \frac{\beta}{R}\int_0^{\infty}\left(\frac{V^2\tau'}{2R}\right)VdV \exp\left(\frac{-\beta\tau' V^2}{2R}\right), \tag{16}$$

$$= \frac{\beta\tau'}{2R^2}\int_0^{\infty} V^3dV \exp\left(\frac{-\beta\tau' V^2}{2R}\right), \tag{17}$$

$$\bar{P} = 1/\beta\tau'. \tag{18}$$

$$\beta = 1/\bar{P}\tau', \tag{19}$$

$$= R/\langle V^2\rangle\tau'. \tag{20}$$

Equation (13) is now rewritten, substituting for β the value just obtained in (20),

$$d\dot{p} = (VdV/\tau'\langle V^2\rangle) \exp(-V^2/2\langle V^2\rangle). \tag{21}$$

A convenient function to obtain experimentally is the average counting rate of the pulses which have a height exceeding the bias setting of a discriminator circuit. If the bias setting is V_0, it follows that the rate $\dot{p}$ should equal

$$\dot{p}_{V_0} = \frac{1}{\tau'\langle V^2\rangle}\int_{V_0}^{\infty} VdV \exp\left(\frac{-V^2}{2\langle V^2\rangle}\right), \tag{22}$$

$$= (1/\tau') \exp(-V_0^2/2\langle V^2\rangle). \tag{23}$$

Results which are essentially the same as (23) can be found in the excellent mathematical analyses of the noise current envelope by Kac[6] and Rice.[7] In order to demonstrate this agreement, it is necessary to apply their results to the number of times per second the noise voltage is expected to pass through V_0 with a positive slope, as seen on the output of a broad band linear amplifier.

Although the Lorentz mean collision time is not directly observable when $\tau' \gtrsim \tau$, the value of τ can be calculated by using (21) together with a random walk solution after Wang and Uhlenbeck.[10] These authors

[10] Ming Chen Wang and G. E. Uhlenbeck, Rev. Mod. Phys. 17, 323 (1945).

have shown that for a one-dimensional model, the probability of a resultant displacement $\nu = n - m$ out of a total of S unit random steps equals

$$P(n|m,S=)\frac{S!}{\left(\frac{\nu+S}{2}\right)!\left(\frac{\nu-S}{2}\right)!}(\tfrac{1}{2})^S. \quad (24)$$

If an assumption is made that $S \gg \nu$, and Sterling's approximation is applied, the following relation is obtained from (24),

$$P_{\nu,S} \propto \exp(-\nu^2/2S). \quad (25)$$

The number of random electron displacements S during an interval of observation τ' is given by

$$S = N\tau'/\tau, \quad (26)$$

where N is the total number of conduction electrons in the resistor. By combining (25) and (26) one obtains the following relation for the probability that ν-electrons will contribute to a pulse during the interval of observation τ'.

$$P_{\nu,\tau'} \propto \exp(-\nu^2\tau/2N\tau'). \quad (27)$$

As (21) and (27) are associated with the same event, it is permissable to equate the exponents of the two expressions. Thus

$$\nu^2\tau/2N\tau' = V^2/2\langle V^2\rangle. \quad (28)$$

It is necessary now to obtain the relationship between the pulse amplitude V and the number of electrons ν which contribute to the pulse in order to evaluate τ in (28).

The pulse amplitude V can be expressed by Ohm's law as the product of the peak noise current I (based on ν electrons) and the constant resistance R (based on the total number of conduction electrons N). Thus the current I and resistance R can be generally expressed as

$$I = \nu\epsilon\lambda/L\tau'\sqrt{3}, \quad (29)$$

and

$$R = L^2/N\epsilon\mu, \quad (30)$$

where ϵ is the electronic charge, λ is the Lorentz mean free path, L is the linear length of the resistor, and μ is the electronic mobility. A few simple steps of algebra applied to (28)–(30) yields

$$\tau = \epsilon\lambda^2 R/(3\langle V^2\rangle\tau'\mu). \quad (31)$$

Analyses of the mean square noise voltage $\langle V^2\rangle$ by Bernamont[2] and Moulin[11] show that a resistor and its distributed capacitance form a system in which the equipartition principle applies. Thus, $\langle V^2\rangle$ can be expressed as

$$\langle V^2\rangle\tau'/2R = C\langle V^2\rangle/2 = kT/2, \quad (32)$$

where C is the distributed capacitance.

[11] E. B. Moulin, *Spontaneous Fluctuations of Voltage* (Oxford University Press, London, 1938).

Substituting (32) into (31) obtains

$$\tau = \epsilon\lambda^2/3\mu kT. \quad (33)$$

It is felt that this derivation of τ gives a clear and convincing model of the Johnson noise. One can quickly obtain the same result by combining the well-known formula $\epsilon/m = \mu/\tau$ with the mean square velocity $\langle v^2\rangle$ from classical statistics,

$$\langle v^2\rangle = (\lambda/\tau)^2 = 3kT/m. \quad (34)$$

Assuming that N is known, the number of electrons ν which contribute to a pulse of a given height can be determined from (27) as

$$\nu = (N\tau' V^2/\tau\langle V^2\rangle)^{\frac{1}{2}}, \quad (35)$$

since τ' can be obtained from the experimental curve of (23) and τ can be obtained from (33). It is not surprising to note that if there were no limitations of amplifier resolving power, and (5) were substituted for (4), then it follows by (35) that

$$\langle \nu^2\rangle = N. \quad (36)$$

Furthermore, assuming still that $\tau' \geqslant \tau$, it follows by (32) that

$$\langle V^2\rangle/R = kT/\tau, \quad (37)$$

which has been derived by Bakker and Heller[3] using Fermi-Dirac statistics. By (37), the mean energy per pulse $\bar{E}$ equals

$$\bar{E} = kT, \quad (38)$$

and as the mean number of electrons involved in the pulse equals $(N)^{\frac{1}{2}}$ by (36), the energy contribution per electron E_0 is given by

$$E_0 = kT/(N)^{\frac{1}{2}} \quad (39)$$

which, in this model, represents the potential drop for an electron falling across a Lorentz mean-free path.

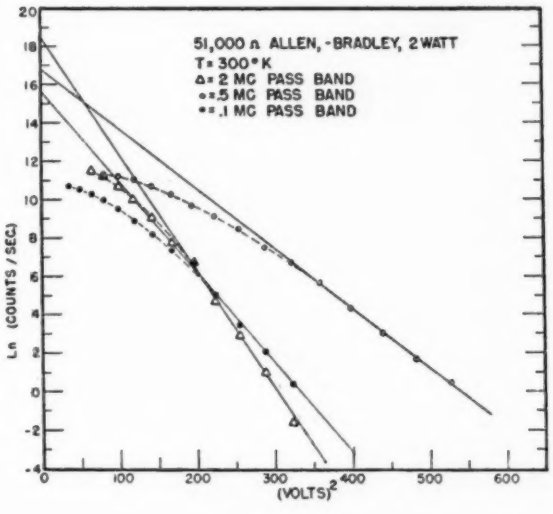


FIG. 3.

MEASUREMENTS

The measurements were taken with the following equipment all of which was made by the Instrument Department of the Oak Ridge National Laboratory: (1) A-1 linear preamplifier, (2) A-1 linear amplifier, (3) 64 scaler, Model Q-762, (4) 1024 scaler, Model Q-834. The A-1 linear preamplifier and amplifier have been described by Jordan and Bell.[12]

Pulse-height measurements taken with different types of commercial grid resistors are shown in Figs. 1–3; and these are the experimental curves which correspond to (23). Curve 5 (Fig. 1) with the input grid grounded gives the shot noise contribution. It can be seen from Figs. 1 and 2 that with the 2-mc pass band of the A-1 amplifier the resolution time of the system τ' extrapolates to $\sim 2.5 \cdot 10^{-8}$ sec. This is the time required by the pulse-height discriminator to receive enough energy to trigger, and consequently it is not inconsistent that it is less than the rise time of the amplifier, i.e., ~ 0.15 μsec. In measuring the random counting rate above $\sim 10^3$ counts/sec. a correction for the recovery time T was made using the equation[13]

$$n_1 = n(1-nT), \quad nT \ll 1,$$

where n_1 is the observed rate, and n is the true rate. The extrapolation of the curves in Figs. 1–3 is justified on the basis of this correction.

In order to predict the curves of Figs. 1–3 from Eq. (23) it is only necessary to have the parameters $\langle V^2 \rangle$ and τ' evaluated. $\langle V^2 \rangle$ can be determined by solving Nyquist's equation for a parallel RC circuit where C is the grid input capacitance, while τ' can be estimated from the design of the equipment.

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to many people who have made valuable suggestions, and with whom he has had stimulating discussions on the subject. Special thanks are indeed due to Mr. P. R. Bell of ORNL and Mr. Keith Henry of NEPA.

[12] W. H. Jordan and P. R. Bell, Rev. Sci. Inst. **18**, 703 (1947).

[13] See W. C. Elmore and M. Sands, *Electronics* (McGraw-Hill Book Company, Inc., New York, 1949), p. 207.

Note on the Use of Reiner's Equation for Recognition of Pseudoplastics from Their Flow Curves

VERA OSMAN KRAKAUER
Interchemical Corporation, New York, New York

(Received March 17, 1950)

A method is described which has been found useful to determine by application of the Reiner equation whether certain flow curves obtained with a concentric cylinder rotational viscometer indicate pseudoplastic or Bingham body types of flow.

THERE has been an increasing use of concentric cylinder rotational viscometers suitable for obtaining complete rate of shear-shearing stress diagrams or flow curves, such as the one described by Green[1] from whose characteristics the rheological behavior of the measured sample may be determined. The several commercial viscometers of this type recently placed on the market are being used in diverse industrial fields, with the likelihood of encountering all types of rheological flow.

There is usually no difficulty in recognizing Newtonian flow from its curve of a straight line passing through the origin (Fig. 1A). Thixotropic plastics can be distinguished by the characteristic hysteresis loops formed between the slope of the up curve and the straight line of the down curve (Fig. 1B), particularly if measured to several top rates of shear,[2] and by the intercept on the shearing stress axis. However, non-thixotropic Bingham bodies, which have an intercept on the shearing stress axis, but then increase linearly with rate of shear (Fig. 1C), and pseudoplastics, whose flow curves show a less than linear shearing stress increase with increasing rate of shear, and which appear to start at the origin and to approach an asymptote (Fig. 1D) sometimes present a difficulty.

For pseudoplastics, the rheologist may be able to fit a straight line through the top points of the flow curve which will intersect the torque axis, with the lower points falling off, and obtain a seeming resemblance to Bingham-body flow curves where the lower points fall off due to plug flow caused by only partial shearing throughout the clearance, due to the yield value. In addition, over the practical range the apparent viscosity, which some investigators calculate as shearing stress divided by rate of shear at each point will continually decrease with increased shear for both types of flow. Therefore, at times, difficulties of recognition will arise. It is important to avoid erroneous conclusions and recognize the true nature of the sample in question that no misleading yield values be calculated for pseudo-

[1] H. Green, Ind. Eng. Chem. (Anal. Ed.) **14**, 576 (1942).

[2] H. Green and R. N. Weltmann, Ind. Eng. Chem. (Anal. Ed.) **15**, 201 (1943).

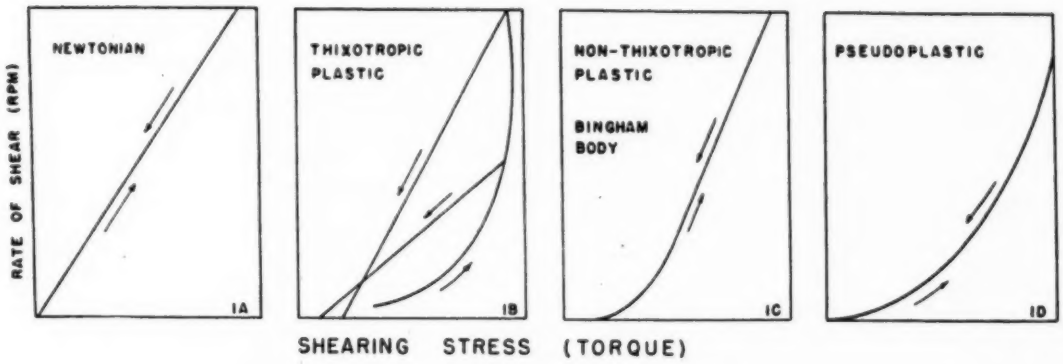


FIG. 1. Schematic flow curves for four types of flow.

plastics, which have none, or that no constant plastic viscosity be calculated for them.

Green points out this difficulty,[3] and suggests recourse to a microscopic method described previously,[4, 5] to help decide which type of flow is being encountered. This note describes a simpler, rapid method, that works well in the majority of cases, and requires merely an examination of the flow curve and a simple calculation.

Reiner and Rivlin, in the development of their equation of flow in concentric cylinder rotational viscometers point out[6] that for materials having a yield value, flow will not start immediately, but will only commence at a torque T_0 where,

$$T_0 = 2\pi r_b^2 hf$$

and completely laminar flow, which is never reached at finite flow rates in a capillary viscometer,[7] will commence in a rotational viscometer at a torque T_2 where

$$T_2 = 2\pi r_c^2 hf,$$

where r_c = radius of the cup (outer cylinder) in cm, r_b = radius of the bob (inner cylinder) in cm, f = yield value in dynes cm, h = immersed height of bob in cm. At T_2, for Bingham bodies the plug flow will be overcome and the straight-line portion of the curve, from whose slope the plastic viscosity, and from whose intercept (T_1) the yield value is calculated, will begin. This means that the individual points above torque T_2 should now fall directly on the straight line drawn through the upper points. Within limits of experimental error, this condition is realized for almost all of the collection of the non-thixotropic Bingham bodies measured at this laboratory over recent years. The measured points begin to fall on the line, at, or slightly above the calculated T_2 (Fig. 2). Thixotropic samples occasionally show larger deviations, possibly due to additional breakdown on the down curve.

However, when a straight line is fitted to the curve obtained for pseudoplastics to a certain top rate of shear, and T_2 is calculated for the yield value obtained from the intercept on the torque axis, it is seen that the individual points do not fall on the line until a much higher torque is reached than that attributable to plug flow if there were a yield value, often as much as 3 and 4 times the calculated value. Particularly when the measurements are continued to higher top rates of shear, it can be seen that the respective straight lines drawn will give materially higher intercepts with increasingly larger deviations from the respective calculated T_2 (Fig. 3), while for non-thixotropic Bingham bodies, measurements to successively higher rates of shear will

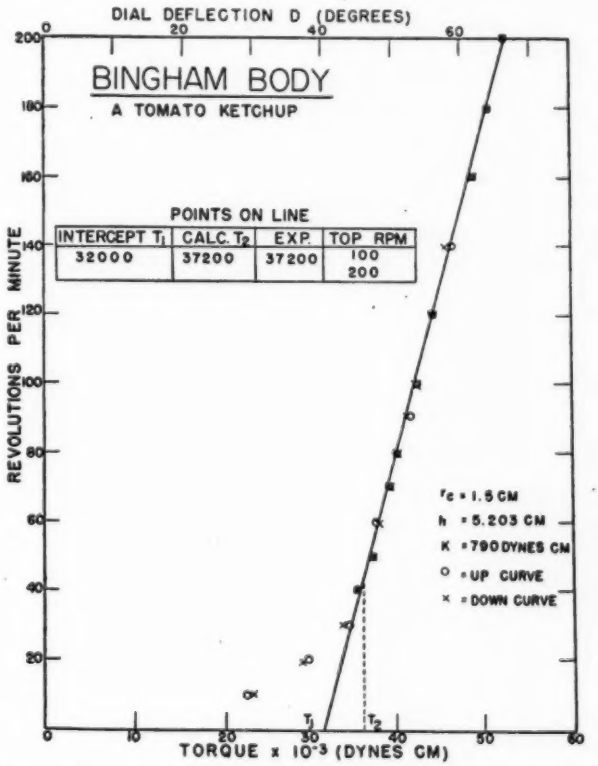


FIG. 2. Experimental flow curve obtained on rotational viscometer for a Bingham body, and calculations.

[3] H. Green, *Industrial Rheology and Rheological Structure* (John Wiley and Sons, Inc., New York, 1949).
[4] H. Green, Proc. A.S.T.M. **20**, Part II, 451 (1920).
[5] H. Green and G. S. Haslam, Ind. Eng. Chem. **17**, 7, 726 (1925).
[6] M. Reiner and R. Rivlin, Kolloid Zeits. **43**, 1 (1927); J. Rheology **1**, 5 (1929).
[7] E. Buckingham, Proc. A.S.T.M. **21**, 1154 (1921).

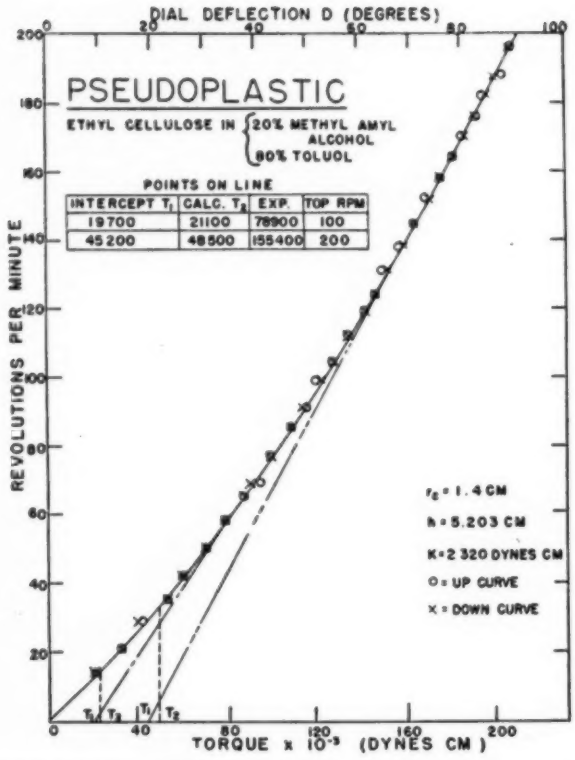

DIAL DEFLECTION D (DEGREES)
PSEUDOPLASTIC
ETHYL CELLULOSE IN 20% METHYL AMYL ALCOHOL 80% TOLUOL
POINTS ON LINE
INTERCEPT T1 | CALC. T2 | EXP. | TOP RPM
19700 | 21100 | 78900 | 100
45200 | 48500 | 155400 | 200
REVOLUTIONS PER MINUTE
TORQUE x 10⁻³ (DYNES CM)
re = 1.4 CM
h = 5.203 CM
K = 2 320 DYNES CM
O = UP CURVE
X = DOWN CURVE

FIG. 3. Experimental flow curve obtained on rotational viscometer for a pseudoplastic, and calculations.

not affect either the yield value or T_2. Simple inspection of the flow curve and the number of points off the line will not usually suffice, since, depending upon the instrumental dimensions, the yield value, and the particular r.p.m. steps chosen, the number of points with too low a torque may vary for a particular Bingham body, while, depending upon the top r.p.m. chosen for the measurement of a pseudoplastic, and the number of points obtained, the amount of curvature may appear to vary. Therefore, it would be preferable to reach sufficiently high rates of shear and to calculate T_2 from the equation. When the flow curves for the McMichael type viscometer are simply plotted directly from the viscometer data as r.p.m. *vs.* angular deflection, changing the deflection values to torque can be avoided by dividing through by the torsion constant (K) of the spring, wire, or other torsion member used,

$$\frac{2\pi r_c^2 hf}{K} = D_2;$$

or, from the equation $f = T_1 C$, which is derived from the Reiner concept, where

$T_1 = D_1 K$,

D_1 = intercept on deflection axis,

$$C = \frac{(1/r_b)^2 - (1/r_c)^2}{4\pi h} \Big/ \ln \frac{r_c}{r_b} = \text{instrumental constant for yield value calculation;}$$

or one can substitute and altogether avoid calculation of f (which would be meaningless for pseudoplastics), using

$$2\pi r_c^2 hCD_1 = D_2,$$

where D_2 is the deflection at which laminar flow ensues and the point should fall on the line. Figures 2 and 3 were plotted from data obtained with a Precision-Interchemical viscometer at increasing and decreasing r.p.m. The coincidence of the "up" and "down" points indicates that no thixotropic breakdown took place. Both graphs could have remained in terms of deflection as indicated along the top of their abscissas. For the Bingham body (Fig. 2), the data for the lines at both top r.p.m. would have been $D_1 = 40.5$, $D_2 = 47$ and D actually on line also 47. For the pseudoplastic curve (Fig. 3), for the 100 top r.p.m. line $D_1 = 8.5$, $D_2 = 9.1$ and D actually on line = 34, and for the 200 top r.p.m. line $D_1 = 19.5$, $D_2 = 20.9$ and D on line = 67. Similarly, for the concentric cylinder modified Stormer type viscometer often used for non-thixotropic samples, whose flow curves are plotted as r.p.m. *vs.* weight where

$T_1 = (W_1 gS/G)$ and W_1 = intercept on weight axis
g = acceleration due to gravity
S = radius of drum to which weight is attached
G = gear ratio between drum and shaft of bob,

one can simply use

$$2\pi r_c^2 hCW_1 = W_2,$$

where W_2 is the weight after which for Bingham bodies the points should fall on the line.

Thus a convenient, rapid method is available to check when doubt exists as to whether a certain viscometric flow curve indicates a pseudoplastic or a Bingham body type of flow. When true yield value is present, the lower points will have too low a torque due to plug flow, but will fall on the straight line at or slightly after the torque at which completely laminar flow ensues, according to the Reiner equation. However, for pseudoplastics, where the intercept is simply a result of fitting a straight line to the upper portion of the curve resulting from the shear-produced reduction of torque, and where both the line and its intercept have no physical significance, the points will usually not fall on the line at or near the calculated value of T_2.

An X-Ray Study of Thermally Induced Stresses in Microconstituents of Aluminum-Silicon Alloys*

JOHN P. NIELSEN† AND WALTER R. HIBBARD, JR.‡

(Received March 13, 1950)

The changing width of high angle Debye x-ray diffraction lines from eutectiferous aluminum-silicon alloys during thermal cycling as measured by a recording Geiger counter spectrometer is interpreted in terms of stresses resulting from the differences in thermal expansion of the microconstituents.

IN a solid alloy containing a dispersed second phase, stresses theoretically should develop in the matrix during cooling from some stress-free state due to the difference between the thermal expansion coefficients of the two phases. If the matrix has the larger coefficient, the stress development process resembles that in shrink or interference fitting except that the components are microconstituents. László[1,2] developed the stress analysis techniques for a variety of such stresses, as they might occur in different ideal microstructures, labeling them "tessellated stresses." Metallographic evidence of the microstructural stresses reported by Boas and Honeycombe[3] consists of slip lines, cracks, and distortion produced by several heating and cooling cycles of certain alloys. The present work is a x-ray diffraction study of such stresses in aluminum-silicon alloys.

Eutectic and certain hypereutectic alloys of this system contain essentially silicon rich particles imbedded in an aluminum rich matrix below the eutectic temperature. Since the thermal expansion coefficient of aluminum is about seven times greater than that of silicon, the aluminum matrix tends to contract against the silicon particles on cooling these alloys, developing a pressure both in the matrix and in the silicon. If, as a first approximation, the silicon particles are considered as uniformly dispersed spheroids, the stress analysis of the matrix becomes that of the pressure on the inner wall of a thick spherical shell, producing radial compression and tangential tension, which depend upon the elastic moduli of the constituents and the dimensions of the shell. Based on physical constants from the 1948 *Metals Handbook* and the equations of László[1,2] and Love,[4] the calculated thermally induced stresses for certain of these alloys are given in Table I. These data indicate that the stresses may reach significant magnitudes and in real alloys where the silicon particles are far from spherical, the stresses at local points should readily exceed the yield strength of the matrix. It is to be noted that the difference between the radial and the tangential stresses at a point acts to produce plastic deformation, with the radial pressure considered as negative in sign.

EXPERIMENTAL PROCEDURE AND RESULTS

The measurement of the width of Debye x-ray diffraction lines appeared to be the best method for obtaining direct evidence of the thermally induced stresses in the aluminum matrix, since both tensile and compressive stresses would tend to shift the line slightly in both directions resulting in broadening. Aluminum-silicon alloys (12, 20, and 30 percent Si) were modified with NaF, chill cast, hot forged to further disperse the silicon particles, and annealed 20 hours at 500°C. A calibrated goniometer and a Geiger tube from a Philips x-ray spectrometer were assembled around a Seemann demountable-target-type x-ray tube so as to cover Bragg angles from 40° to 83°. By using a selected beam from a line focus source[5] fine slits of 0.005 in. at the source and at the Geiger tube permitted sufficient intensity for the delineation of line breadths in the back-reflection region. In order to take full advantage of the x-ray focusing principle,[6] the specimen was mounted so that the effective diffraction surface could be experimentally adjusted tangent to the focusing circle. To further enhance resolution, the scanning speed of the Philips goniometer was reduced to 0.05° Bragg per minute. Relative x-ray intensities were recorded with a chart speed such that one degree Bragg required 40 in. of chart. Line breadth measurements were made at several temperatures on cooling from 400°C to room temperature using the 420 diffraction line of the aluminum matrix and Cu$K\alpha$-radiation. The temperatures were maintained by an electrical resistance furnace which fitted over the specimen holder and contained narrow slits for the exit and entry of the beam. The temperature

* From a dissertation presented by John P. Nielsen to the Graduate School of Yale University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

† Associate Professor of Metal Science, New York University, New York, New York.

‡ Assistant Professor of Metallurgy, Yale University, New Haven, Connecticut.

[1] F. László, J. Iron and Steel Inst., Part I, No. 1, 173P (1943); Part II, No. 2, 137 (1943); Part III, No. 2, 183P (1944); Part IV, No. 2, 207P (1945); Part V, **164**, 5 (1950).

[2] F. R. N. Nabarro, László's Papers on Tessellated Stresses, Symposium on Internal Stresses in Metals and Alloys, Inst. of Metals, London (1948).

[3] W. Boas and R. W. K. Honeycombe, Proc. Roy. Soc. **186**, 57 (1946); **188**, 427 (1947). J. Inst. Metals (January, 1947). Council for Sci. and Ind. Res., Melbourne, Australia, Serial No. A. 151, Physical Metallurgy Report No. 4 (May 6, 1947). Nature, **153**, 494 (1944); **154**, 388 (1944).

[4] A. E. H. Love, *Mathematical Theory of Elasticity* (Cambridge University Press, London, 1920); third edition, p. 140.

[5] N. C. Baese, Rev. Sci. Inst. **8**, 258 (1937).

[6] J. C. M. Brentano, Proc. Phys. Soc. **37**, 184 (1925).

TABLE I. Calculated microstructure shrink-fit stresses developed on cooling from 300°C to different temperatures in aluminum shell-silicon core compound sphere. Stresses in 1000 p.s.i.

Wt. percent Si	To: 200°C	Tangential stress (tension) Inner surface 100°C	20°C	200°C	Outer surface 100°C	20°C	Radial stress (compression) Inner surface[a] 200°C	100°C	20°C
12	14	29	41	4	8	11	20	42	60
20	14	29	41	4	8	11	20	42	60
30	17	35	49	8	17	24	17	35	49
40	19	39	55	12	25	35	13	27	38

[a] Radial stress at outer surface of aluminum shell vanishes.

was measured by a thermocouple placed in a hole drilled in each specimen and was controlled to within ±2°C. The deviation in line breadth produced by the 2°C variation was negligible. The results are tabulated in Table II.

TABLE II. Widths in radians, W_r, at half-maximum for the 420 Cu$K\alpha_1$-aluminum line from aluminum-silicon alloys at various temperatures (multiply by 10^{-3}).

0 percent Si		12 percent		20 percent		30 percent	
T°C	W_r	T°C	W_r	T°C	W_r	T°C	W_r
260	3.5	420	3.5	307	1.6	313	3.5
22	2.6	320	3.5	272	2.5	260	3.2
		262	3.5	214	3.2	206	5.1
		211	3.0	182	3.5	153	5.1
		142	3.3	144	3.2	96	5.4
		77	4.2	97	3.3	24	5.6
		12	4.0	31	3.7	360	4.3
				22	3.9	22	5.2

A metallographic confirmation of the stresses was attempted¶ similar to the Boas and Honeycombe experiments. It was found that polished specimens of these alloys would not develop slip lines[7] on thermal cycling, but neither could slip lines be developed by mechanical straining (squeezing in a vise), so that it does not appear possible to obtain metallographic support for the x-ray data.

EVALUATION OF RESULTS

As expected, cooling the aluminum (0 percent Si) specimen from 260°C to room temperature produced no line broadening, in fact, the high resolution even revealed a line sharpening. However, all alloys containing added silicon showed line broadening on cooling, as predicted by the stress analysis. No pronounced asymmetry was found in the shape of the lines, indicating no measureable difference in the magnitude of the tensile and compressive components. On reheating the 30 percent silicon specimen, it was found that the line width decreased somewhat, lending support to the thesis that elastic stresses were being detected. The gradient nature of the stresses together with the fact that the silicon particles were not spherical precluded any quantitative correlation with the calculated stresses. Furthermore, the x-ray measurements involved only the region at the surface of the specimen. Regarding the solid solubility of silicon in the aluminum phase, calculations indicate that the 0.05 percent to 0.48 percent solubility limits between room temperature and 400°C would disturb the x-ray lines less than the probable error of measurement. However, the gradient nature of the stresses and any cold working resulting from stresses exceeding the yield point of the matrix might have added to the line broadening on cooling.

¶ Experiment performed by W. H. Hawley, Jr., graduate student at Hammond Laboratory, Yale University.

[7] A "rumpled" surface indicating strain was visible to the naked eye, but, under the microscope, fine dark lines indicative of slip were not visible. For a similar phenomenon see Guy, Barrett, and Mehl, Trans. A.I.M.E. **175**, 216 (1948).

SUMMARY

Thermally induced microstructure shrink-fit stresses produced in the aluminum matrix around silicon particles in duplex aluminum-silicon alloys are analyzed on the basis of elastic theory, and x-ray diffraction experimental evidence is presented to confirm their presence.

ACKNOWLEDGMENT

Acknowledgment is due the North American Philips Company, New York, New York for their generous loan of the Geiger counter spectrometer. The personal efforts of Dr. O. S. Duffendack, Director of Research, Dr. W. Parrish, and Mr. C. J. Woods, of that company in arranging for the loan are particularly appreciated. Dr. Kent R. Van Horn of the Aluminum Company of America, Cleveland, Ohio, arranged for the supply of specimen material. Professor L. W. McKeehan, Director of Sloane Physics Laboratory, Yale University and Dr. C. H. Mathewson Department of Metallurgy, Yale University kindly read the manuscript and offered helpful suggestions.

X-Ray Diffraction by Elastically Deformed Crystals*

J. E. White
*Acoustics Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts***
(Received March 20, 1950)

Published articles establish the fact that vibration or other inhomogeneous strain greatly increases the ability of a crystal to diffract x-rays. The present article gives experimentally determined relations between integrated intensity and strain for the rather simple case of static bending. All measurements were made on quartz plates, but a wide range of geometrical and crystalline parameters was covered nevertheless. In some cases, the integrated intensity for extreme bending was more than 20 times that from the unbent crystal. An approximate theory is developed which agrees well enough with the measurements to serve as a means of computing the effects of other types of strain or of other combinations of crystalline parameters. A qualitative study of vibrating crystalline plates is also described briefly.

NOMENCLATURE

e	electronic charge
m	electron mass
c	velocity of light
N	number of unit cells per unit volume
λ	x-ray wave-length
F	crystalline structure factor
θ	Bragg angle
ϕ	variable angle of incidence
l	distance traversed in crystal
R	radius of bending
$2s$	angular width for perfect reflection
w	"average" width
ϵ	extinction coefficient
ϵ_{ave}	average extinction coefficient
μ	linear absorption coefficient
A_0	integrated intensity for perfect crystal
B, L	integrated intensities
t	thickness of crystal
Θ	angle between crystalline plane and face of crystal
I	beam intensity
P_0	beam power
ω	angular velocity of crystal
E	energy diffracted in one traversal

INTRODUCTION

THE influence of elastic vibrations of a crystal on the diffraction of x-rays by the crystal has interested many investigators during the last twenty years.[1] One fact clearly established by their work is that Laue spots formed when an x-ray beam traverses a crystal vibrating in some resonance condition are much more intense than they are when the crystal is at rest. A similar increase for Bragg reflection off the surface was found. Statically bent quartz was also shown to exhibit this effect. The increased intensity was generally attributed to a reduction in extinction, either primary or secondary. A theory[2] treating diffraction by a crystal whose planar spacing varies sinusoidally in space and time was published, but extinction effects were not considered, and as a result, the integrated intensity of any reflection came out to be independent of vibration.

The principal objective of the work being described here is the determination of a quantitative relationship between increased diffracting power and degree of deformation. Since high frequency vibration patterns may be quite complex, static bending was chosen as a type of strain which can be realized experimentally and described easily. Quartz was chosen for study because of its excellent elastic properties and the high degree of crystalline perfection obtainable. The experimental procedure and results will follow the discussion of the approximate theory, since the terminology and basic objectives can best be presented in connection with the theory. The expressions derived provide a means of generalizing the results to other crystals and other types of strain, including some forms of vibration. No measurements were made on vibrating crystals, although the qualitative experiments performed provide further evidence that it is the local curvature of the crystal which controls the increase in diffracting power.

AN APPROXIMATE THEORY

The mathematical description of x-ray diffraction by a distorted crystal as developed here relies upon the Darwin[3] treatment of x-ray diffraction by a perfect crystal for fundamental constants required. He found that a plane monochromatic beam is perfectly reflected over a small angular range which can be expressed (for σ-polarization),

$$2s=(2e^2N\lambda^2F)/(mc^2\pi\sin 2\theta). \tag{1}$$

Outside this range, the reflected amplitude drops sharply. Even so, the area under the "tails" is one-third the area under the flat portion of the curve. The area under this curve is customarily expressed in units applicable to an experiment in which the crystal is rotated at a uniform angular velocity ω and the total energy E is measured as the region of reflection is traversed once. Expressed as a ratio to the incident power P_0,

$$\frac{E\omega}{P_0}=A_0=\frac{8N\lambda^2Fe^2}{3\pi\sin 2\theta mc^2}\frac{1+\cos 2\theta}{2}. \tag{2}$$

* Extract from a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the Massachusetts Institute of Technology.

** Now at the Magnolia Petroleum Company, Field Research Laboratories, Dallas, Texas.

[1] W. G. Cady, *Piezoelectricity* (McGraw-Hill Book Company, Inc., New York, 1946).

[2] J. Weigle and K. Bleuler, Helv. Phys. Acta **15**, 445 (1942).

[3] C. G. Darwin, Phil. Mag. **27**, 315 and 675 (1914).

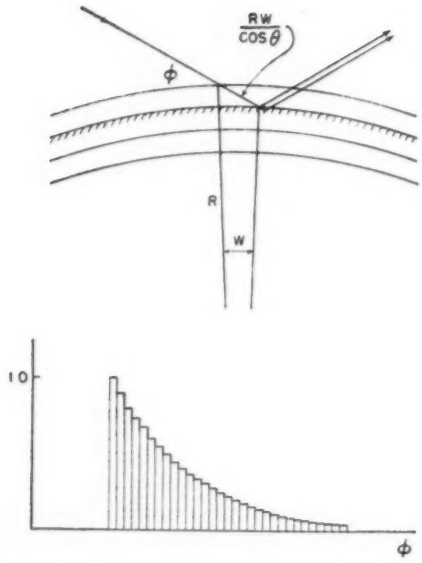


FIG. 1. Geometry for approximate theory (above) and resulting rocking curve (below).

This is the integrated intensity for a perfect crystal when the incident beam is unpolarized.

Darwin also found that the transmitted beam is rapidly "extinguished" within this angular range. The reduction in amplitude as the beam penetrates the crystal along a path l can be expressed as an exponential decay,

$$I_l = I_0 e^{-\epsilon l},$$

where

$$\epsilon = (2\pi \sin 2\theta)[s^2 - (\phi - \theta)^2]^{\frac{1}{2}}/\lambda. \tag{3}$$

Within the region of perfect reflection, the extinction coefficient ϵ varies, being a maximum at the middle and zero at the edges. Outside the region of perfect reflection, ϵ is imaginary and therefore does not represent attenuation. The values of ϵ are such that reflection by a perfect crystal is practically a surface phenomenon.

The essential feature of the theory described here is the assumption that a distorted crystal behaves locally like a perfect crystal, the strains becoming apparent only when viewed at points many wave-lengths apart. It is

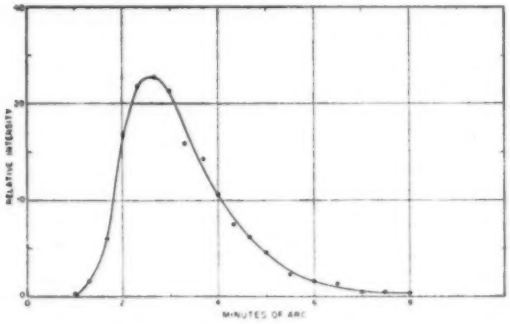


FIG. 2. Rocking curve for bent crystal.

developed for a crystal bent into a section of a cylinder of radius R. An expression for the area under the rocking curve as a function of bending is our objective. Before proceeding, let us make two simplifying approximations regarding the behavior of a perfect crystal. The first is that perfect reflection occurs over a range w such that the area A_0 is the same as that given by the Darwin treatment. This makes $w = 2.66s$ as compared with $2s$. The second assumption is that the extinction coefficient ϵ, which governs the depth of penetration of the beam before it is reflected, has the constant value ϵ_{ave} over the whole range w and is zero elsewhere. ϵ_{ave} is the linear average of Darwin's ϵ over the angular range w, and it is given by

$$\epsilon_{Av} = \frac{3\pi^2 \sin 2\theta s}{8\lambda} = \frac{3\pi e^2 N \lambda F}{8mc^2}. \tag{4}$$

We are now ready to state how the integrated intensity from a bent crystal should depend on the radius of bending. As a bent crystal (Fig. 1) is rotated counter-

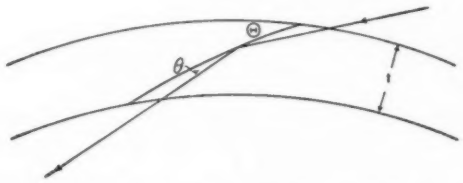


FIG. 3. Geometry for plane not parallel to surface of crystal.

clockwise, the conditions for reflection of the elementary pencil of x-rays shown will become satisfied for the region near the surface. Throughout a range w, reflection is perfect, and a contribution of A_0 is made to the integrated intensity. The beam penetrates only a very short distance controlled by ϵ_{ave}, and the rest of the crystal is shielded. As the angle proceeds beyond w, however, the beam is no longer reflected at the surface, but can penetrate with attenuation due only to ordinary absorption. After penetrating a distance $Rw/\cos\theta$, it reaches a region which has been bent through w with respect to the surface, counteracting the rotation w already given to the whole crystal, so conditions are right for reflection. (Since the whole angular range for reflection is very small, the angle can be considered constant at the Bragg angle θ for purposes of geometry.) During the next w radians, A_0 is contributed by the shaded area in Fig. 1, but the beam is reduced by $\exp[-(2\mu Rw/\cos\theta)]$ in twice traversing the surface layer, where μ is the linear absorption coefficient. In this fashion, the crystal is considered made up of discrete layers, each reflecting like a perfect crystal. This proceeds until the back side of the crystal is properly oriented, or until the total distance is $t/\sin\theta$, t being the thickness of the crystal. This process can be stated as a

summation,

$$B_0 = A_0 \sum_{n=0}^{\frac{t\cos\theta}{Rw\sin\theta}} \exp[-n2\mu Rw/\cos\theta],$$

$$B_0 = A_0[1-\exp(-2\mu t/\sin\theta)]/[1-\exp(-2\mu Rw/\cos\theta)].$$

It is apparent that as bending is increased, the layers described above become thinner, whereas the distance required for almost complete reflection remains fixed. In Fig. 1, the shaded area becomes deeper than a layer as bending becomes extreme. This "thin crystal" effect reduces the contribution of each layer and causes the curve of integrated intensity *versus* bending to flatten off. A statement of this effect was derived in the following way. In a thick perfect crystal, the transmitted beam is reduced exponentially, as derived by Darwin. Let us state that this is true for our crystal, and that ϵ_{ave} is the proper exponent. Thus $dI/I = -\epsilon_{ave}dl$. Let us further state that the rapid loss of energy from the transmitted

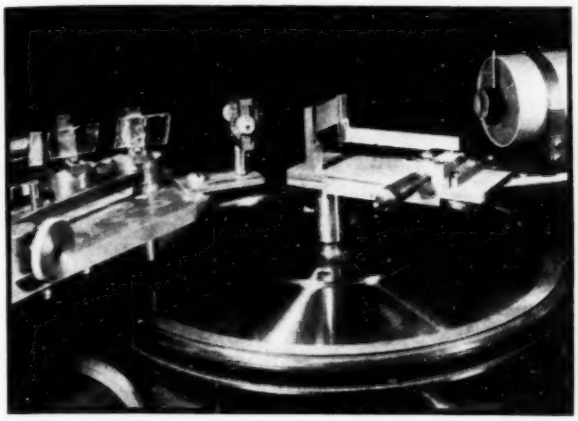


FIG. 4. Apparatus including monochromator, bent crystal, and counter.

beam is due to its transfer to the reflected beam, so that the rate of increase of the reflected beam equals the rate of loss from the transmitted beam.

$$dI_R = -dI = I_0 \exp(-\epsilon_{ave}l)\epsilon_{ave}dl,$$

$$I_R = I_0 \int_0^{Rw/\cos\theta} \exp(-\epsilon_{ave}l)\epsilon_{ave}dl$$

$$= I_0[1-\exp(-\mathrm{RW}\epsilon_{ave}/\cos\theta)].$$

I_R/I_0 is the factor by which reflection is reduced when a crystal becomes thin. Hence, a complete expression for integrated intensity from a bent crystal is

$$B = A_0[1-\exp(-2\mu t/\sin\theta)] \times[1-\exp(-\epsilon_{ave}wR/\cos\theta)]/[1-\exp(-2\mu wR/\cos\theta)]. \quad (5)$$

The step-wise procedure described above would lead to a rocking curve as sketched at the bottom of Fig. 1. A

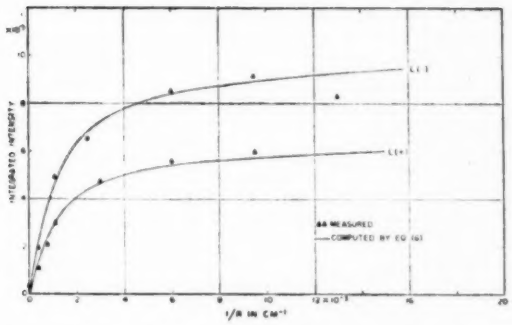


FIG. 5. Integrated intensity for 20.0 plane at 38°13′ with crystal face using Mo$K\alpha$.

highly monochromatic beam extremely limited in divergence and area gave the experimental curve in Fig. 2, which bears a strong resemblence to the shape suggested. The rocking curve obtained for no bending was narrower and symmetrical, and bending in the opposite direction gave correspondingly reversed asymmetry.

Although symmetrical Bragg reflection probably represents the most frequent usage of a crystal, it is nevertheless a special case, and it seemed desirable to work out an expression similar to the one above for diffraction by crystalline planes making an arbitrary angle with the surface of the crystal. The treatment is limited to crystalline planes which are parallel to the axis of bending. The geometry shown in Fig. 3 brings out the fact that an interchange of source and receiver changes the angle between the incident beam and the surface of the crystal. It turns out that the two cases lead to different expressions for integrated intensity, a result which may at first seem surprising. If the diffracted beam emerges on the same side of the crystal as the incident beam, corresponding to $\Theta < \theta$, it is referred to here as a Bragg reflected and designated by B. Similarly, if the beam traverses the crystal, corresponding to $\Theta > \theta$, it is called a Laue reflected and designated by L. In each case, plus and minus signs are added to differentiate the two possible angles of incidence. The steps leading to the expressions given below will not be in-

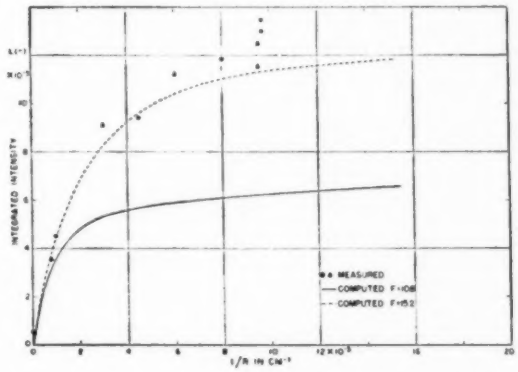


FIG. 6. Integrated intensity for 10.0 plane at 38°13′ with crystal face, using Mo$K\alpha$-radiation.

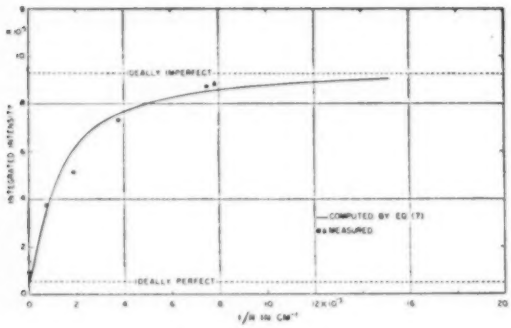


FIG. 7. Integrated intensity for 20.2 plane parallel to crystal face, using Mo$K\alpha$-radiation.

cluded, since the more involved geometry makes the derivation rather lengthy without adding anything to the viewpoint used in the simpler derivation.

$$L(\pm)=A_0\exp[-\mu t/\sin(\Theta\mp\theta)]\times\{1-\exp[\pm\mu t(\sin(\Theta+\theta)-\sin(\Theta-\theta))/\sin(\Theta+\theta)\sin(\Theta-\theta)]\}\times\{1-\exp[-\epsilon_{ave}wR/\cos\theta\cos\Theta(1+\sin^2\Theta)]\}/\{1-\exp[\pm\mu wR(\sin(\Theta+\theta)-\sin(\Theta-\theta))/\sin(\Theta\mp\theta)\cos\theta\cos\Theta(1+\sin^2\Theta)]\}. \quad (6)$$

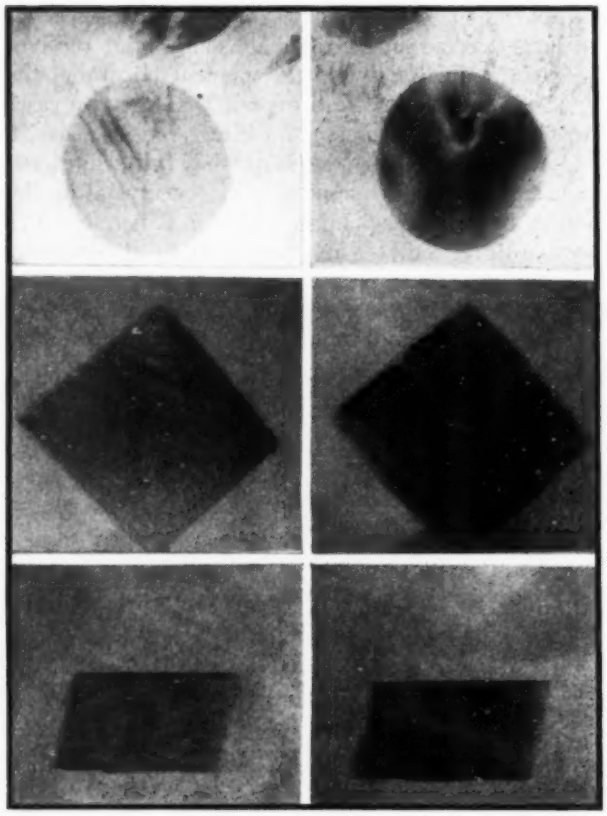


FIG. 8. Qualitative comparison of vibrating crystals (right) with the same crystals at rest (left).

$$B(\pm)=A_0\{1-\exp[-\mu t(\sin(\theta+\Theta)+\sin(\theta-\Theta))/\sin(\theta+\Theta)\sin(\theta-\Theta)]\}\{1-\exp[-\epsilon_{ave}wR/\cos\theta\cos\Theta(1+\sin^2\Theta)]\}/\{1-\exp[-\mu wR(\sin(\theta+\Theta)+\sin(\theta-\Theta))/\div\sin(\theta\pm\Theta)\cos\theta\cos\Theta(1+\sin^2\Theta)]\}. \quad (7)$$

In either case,

$$t\cos\theta\cos\Theta(1+\sin^2\Theta)/Rw\sin(\theta\pm\Theta)>5.$$

These are the expressions which have been compared with experiments on bent quartz. The assumptions made appear rather extreme. In particular, considering the contributions from different depths as a step-wise process is not justified unless there are many contributions. This means that the expressions are not applicable if the bending is too small, as indicated by the quantity following the expressions for integrated intensity. Setting this quantity to be greater than 5 rather than some other integer is arbitrary.

EXPERIMENTS WITH STATICALLY BENT QUARTZ

The apparatus for observing bent quartz is shown in Fig. 4. The beam strikes two calcite crystals set up as a double crystal spectrometer in the antiparallel position, thus achieving a wave-length spread in the final beam even narrower than the natural width of the α_1-line. A slit limits vertical divergence and beam area. The quartz slab is mounted on the axis of a precision turn-table in such a manner that motion of an ordinary micrometer at the end of a lever results primarily in a bending moment. Radius of bending computed from the lever position checked determinations by two other methods. The diffracted beam entered a Geiger tube which was mounted on a second turn-table geared to turn at twice the angular velocity of the first. Readings were taken at closely spaced angular positions by registering the number of counts in an accurately controlled minute by means of a scaling circuit. A plot of these readings, in arbitrary power units against angle in radians, constitutes a rocking curve. The power of the direct beam in the same units was measured by the Geiger tube and counter, although it was necessary to introduce a known absorption in the beam or effect a known reduction in tube output of a factor of 10 or so. The area under each rocking curve was divided by the incident power to obtain integrated intensity. Convex and concave bending were plotted together, and in agreement with the theory, have the same effect on integrated intensity. Measurements were made on two crystals with faces parallel to the 10.1 planes, and a third crystal with faces parallel to the 10.0 planes. In each case, the crystal was lapped to a thickness of about 0.05 cm and then etched for about two hours in 48 percent hydrofluoric acid. Fourteen curves were measured and computed, involving six crystallographic planes, some parallel to the crystal face and some not. Most of the curves were obtained with Mo$K\alpha_1$-radiation, the others with Cu$K\alpha$.

Agreement between computation and measurement was generally good, five of the curves being comparable with the two shown in Fig. 5. These curves illustrate the effect of interchanging source and receiver. Real discrepancies were found for all three curves involving the 10.0 plane, of which Fig. 6 is an example. The solid curve is based on the published structure factor.[4] The value of 10.8 for the 10.0 plane appears to be in error. A value of 16.3 brings computed curves into agreement with all three measured curves, and a computation based on the published structure yielded 15.2, suggesting that an error was made in the original calculation. The dashed curve is computed using 15.2 as the structure factor.

It is reasonable to suppose that the integrated intensity for any crystal lies between the value for perfection and the value resulting from such complete imperfection that extinction is of no consequence.[5] The former limit is given as Eq. (2). The integrated intensity for the ideally imperfect condition is

$$\frac{E\omega}{P_0}=\frac{N^2F^2\lambda^3e^4}{2\mu m^2c^4\sin 2\theta}\,\frac{(1+\cos^2 2\theta)}{2}. \tag{8}$$

These limits are shown as dashed lines in Fig. 7. The measured values almost cover the extreme range, and the computed curve is seen to approach a limit very near the ideally imperfect value.

EXPERIMENTS WITH VIBRATING CRYSTALS

Thin crystalline slabs of quartz and calcite were driven at frequencies of flexural resonance by means of a slender piezoelectric crystal cemented to the back faces of the slabs. The front faces were illuminated by a divergent beam of x-rays from a molybdenum target. The diffracted x-rays were observed on a film set parallel to the crystal and fairly near to the crystal. The crystal was rotated through an angular range such that each point had an opportunity to diffract the $K\alpha$-doublet. The film was mounted on an arm from a turn-table rotating at twice the angular velocity given the crystal. This procedure "scanned" the face of the crystal and gave an approximate picture. In each case, the crystals were etched before use.

Figure 8 shows three pairs of such pictures, illustrating the effect of vibration. The exposure times and photographic histories of both members of each pair are approximately the same, so the increased blackening observed is due to vibration. Other pictures of this kind pretty well established that radius of curvature is the parameter which controls this increased blackening. A more detailed description of the vibration studies is being submitted to the *Journal of the Acoustical Society of America*.

[4] P'ei-Hsiu Wei, Zeits. f. Krist. **92**, 355 (1935).
[5] W. A. Wooster and G. L. MacDonald, Acta Cryst. **1**, 49 (1948).

SUMMARY

The measurements given here show the increase in diffraction due to bending to be a gross effect, even a factor of 20 or more. It seems likely that this mechanism must contribute to the performance of bent crystal monochromators. The theory, although admittedly approximate, provides a means of predicting the effect of bending on various crystals whose crystal constants are known. The same procedure can be used to derive expressions for types of strain other than bending. The computed curve may also be used to compute a time average of diffracted energy when the radius is varying sinusoidally with time, as in flexural vibration.

ACKNOWLEDGMENT

The author wishes to express his appreciation for the suggestions and encouragement so generously given by Professor B. E. Warren during the course of this research and to thank Dr. W. Parrish of the Philips Laboratories for donating the quartz crystals used.

Some Effects of Nozzle Design on the Diffraction of Electrons by Gases

R. B. Harvey, F. A. Keidel, and S. H. Bauer
Department of Chemistry, Cornell University, Ithaca, New York
(Received March 20, 1950)

An evaluation has been made of some of the errors introduced into the electron diffraction of gases by failure of the sample to be localized, and by secondary and higher order scattering.

The degree to which poor nozzle design may contribute to these factors was investigated experimentally by measurements of the pressure in the camera and absorption measurements of the main electron beam when a small volume of gas was allowed to expand into the camera; and theoretically for the cumulative effects of higher order scattering on the final picture. This evidence was complemented by density measurements on photographic plates taken under similar conditions. The effect of using liquid air cooled surfaces around the nozzle was tested.

Of the nozzles tested, better "image to background" ratios were obtained with the "hydrodynamic" and "comparison" nozzles than with the "conventional" nozzle, but the gas must be admitted slowly enough so that less than about 10 percent of the main beam is appreciably scattered. The "image to background" ratios were always higher using adequate cooling on the condensing surfaces surrounding the nozzle. In the case of pressure and electron beam absorption measurements short exposures were advantageous, where secondary scattering was not directly considered, and high camera pumping speeds were always advantageous.

On the basis of these results some general design suggestions for electron diffraction cameras, and requirements for good efficiency in nozzles can be made. A nozzle design is proposed which would reduce materially the errors normally encountered in electron diffraction work.

INTRODUCTION

WHEN a stream of high velocity electrons is scattered by a gaseous sample, the intensity of the scattered radiation may be written:[1]

$$I_s = J_0\left[\frac{8\pi^2 me^2}{h^2}\right]^2 \frac{N}{R^2 s^4}[\text{atomic contribution} + \text{molecular contribution}], \quad (1)$$

where I_s can be defined in terms of the number of electrons crossing unit area per unit time at the focal plane of the diffracted ray; J_0 is the total intensity in the primary beam; N is the total number of molecules intercepted by the beam per unit beam area; $s=(4\pi/\lambda)\times\sin(\theta/2)$ where λ is the electron equivalent wavelength and θ is the angle between the primary and scattered rays; R is the distance from the sample to the focal plane; and $8\pi^2 me^2/h^2 = 2/A_H$ where A_H is the "radius" of the hydrogen atom, and m, e, and h are universal constants. The "atomic contribution" factor is dependent only on the scattering functions of the individual atoms of the molecule, and so accounts for the structure independent contribution to the total scattered intensity, whereas the "molecular contribution" factor is in addition dependent on the interatomic distances in the molecule and so accounts for the structure dependent contribution to the scattered intensity.

The following are some of the assumptions made in the derivation of the above equation:

(1) The N molecules scatter independently with no mutual phase relations. This is satisfied for a gaseous sample.

(2) The sample is of infinitesimal thickness; diffraction occurs at one value of R only. This is practically never true for a gaseous sample.

(3) There is no secondary or multiple scattering (i.e., no scattered electrons are rescattered) in the sample region or in the space between the sample and the detector. This is generally not true for diffraction by a gaseous or a solid sample.

(4) The diffracted beam is perfectly focused at the plane of the detector. This is only approximately true.

A theoretical discussion of the errors which result when these assumptions are invalid appeared in a report[2] written for the Office of Naval Research.

The effects of various gas distributions about the point of gas injection were treated from the standpoint of image resolution, and it was found that indeed a loss of resolution was experienced, the loss becoming increasingly serious with large scattering angles, particularly for the high frequency terms. With regard to the use of a beam trap close to the nozzle to remove the main beam and to reduce the observed scattered intensity at small angles preferentially, so that the density range of the photograph may be reduced and excessive halation prevented, it was found that the positions of the inner maxima and minima of the pattern were not only subject to a similar loss of resolution but also to some lateral shifting of their positions. Experimentally it is also found that the use of a beam trap close to the nozzle leads to further difficulties insofar as edge scattering contributes appreciably to the general background.

A mathematical analysis of multiple scattering within the nozzle indicated that secondary scattering expressed as the ratio of [gain in background/loss in diffracted intensity] is dependent on the diffracting angle, although the fractional loss of any diffracted ray is probably independent of angle. The extent of multiple scattering was found to be independent of the sample distribution along the direction of the electron

[1] P. J. W. Debye, J. Chem. Phys. **9**, 55 (1941).

[2] Bauer, Keidel, and Harvey, "An evaluation of quantitative procedures for the estimation of intensities of diffracted electrons," a report issued on work done under Contract No. N6 ori-213, Task Order I, NR 052-040. This report is available from the Library of Congress, Washington, D. C.

beam but to be proportional to the sample density at the point of diffraction.

The treatment of the last assumption, that pertaining to the effects of imperfect focusing at the plane of the detector, showed that although there was no appreciable shifting of the peaks or valleys of the intensity curves, there was a loss of resolution (the ratio of observed peak heights to the theoretical peak heights was less than 1), and this loss increased with increasing distance for any atomic pair. A slight additional loss in resolution which has angular dependence is also introduced by the use of a flat detector (e.g., a photographic plate) instead of one curved to fit the focal circle.

Although Eq. (1) has been checked quantitatively, within rather broad limits of experimental error,[3] the cumulative effects of all of the specific errors listed above must be carefully evaluated before one may determine precisely the structures of molecules which must be formulated in terms of more than two structural parameters.

The extent of invalidity of assumptions two and three is particularly influenced by nozzle design. It is from this viewpoint that the subject of nozzle design is discussed in the present paper. Some aspects of this study also appeared in the previously mentioned ONR report.[2]

NOZZLE EFFICIENCY DEFINITIONS

Experimentally, in order to obtain a gas diffraction pattern, a jet or stream of gas is allowed to intercept the electron beam. It is therefore apparent that the distance along which electrons are scattered cannot be smaller than that dimension of the gas exit which is parallel to the direction of the electron beam. Actually the gas never leaves the nozzle in a perfectly collimated beam but rather tends to diffuse throughout the camera. Thus, contrary to assumption (2), one obtains, in practice, a superposition of patterns originating at various distances R, rather than a single pattern originating only at R_0, the distance of the nozzle axis to the plane of the detector. Maximum resolution will be obtained only if the spread in R over which appreciable scattering takes place is as small as possible, i.e., when the function $N(R)$ is negligibly small except in the vicinity of R_0. However, a nozzle having a thin slit perpendicular to the electron beam will not be satisfactory in this respect because the issuing gas stream is then poorly collimated. In any case, an estimate of the amount of extraneous scattering could be obtained if the distribution of the gas in the camera could be plotted. It appeared reasonable that if the gas stream intersecting the electron beam directly over the nozzle gave rise to the principal diffraction pattern, then this might be called "useful diffraction," while the gas existing throughout the camera volume would give rise to many overlapping patterns, thus constituting a background, and this diffraction might be termed "harmful." One measure of the ratio of the "useful diffraction" to the total diffraction is then

$$E_p(t)=\frac{Q_{\text{nozzle}}}{Q_{\text{nozzle}}+Q_{\text{camera}}},$$

where Q_{nozzle} is the fraction of the total available sample that is in the electron beam directly over the nozzle, averaged up to time t; and Q_{camera} is the fraction of the total available sample that is in the rest of the electron beam (considered as a cylinder) averaged up to time t. One sees from reference to Fig. 1 that the ratio of the "useful diffraction" to the total diffraction is actually equal to

$$\frac{C_1+C_3}{C_1+C_2+C_3+C_4},$$

where C_n is the relative contribution of process n in Fig. 1, to the total intensity. Therefore a high efficiency $E_p(t)$ indicates that relatively little extension of the sample occurs beyond the dimensions of the nozzle and hence that the loss of resolution due to sample extension alone is small, regardless of how much resolution is lost as a result of multiple scattering (process 3).

We shall also discuss another measurement of efficiency:

$$E_{ab}(t)=\frac{C_1+C_3+C_4}{C_1+C_2+C_3+C_4},$$

which may be estimated from data on the absorption of the main electron beam. Values of $E_{ab}(t)$ should agree fairly well with corresponding values of $E_p(t)$ because process 4 contributes only slightly to the total scattered intensity. It is convenient to distinguish (2a), that part of process 2 taking place inside the chimneys surrounding the nozzle from (2b), that part taking place outside the chimneys. Then, if no liquid air cooling is used, we may write:

$$E_{ab}\text{ (no liquid air)}=\frac{C_1+C_3+C_4}{C_1+C_{2a}+C_{2b}+C_3+C_4},$$

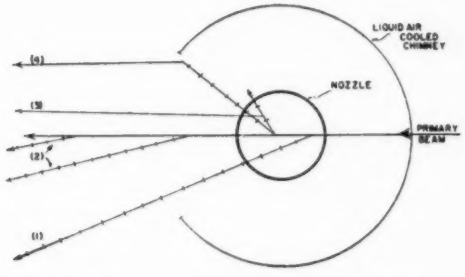


FIG. 1. Schematic representation of various scattering processes (1) Primary diffracted ray—within nozzle. (2) Primary diffracted ray—outside nozzle. (3) Secondary diffracted ray—within nozzle. (4) Secondary diffracted ray—outside nozzle.

[3] J. M. Hastings and S. H. Bauer, J. Chem. Phys. 18, 13 (1950).

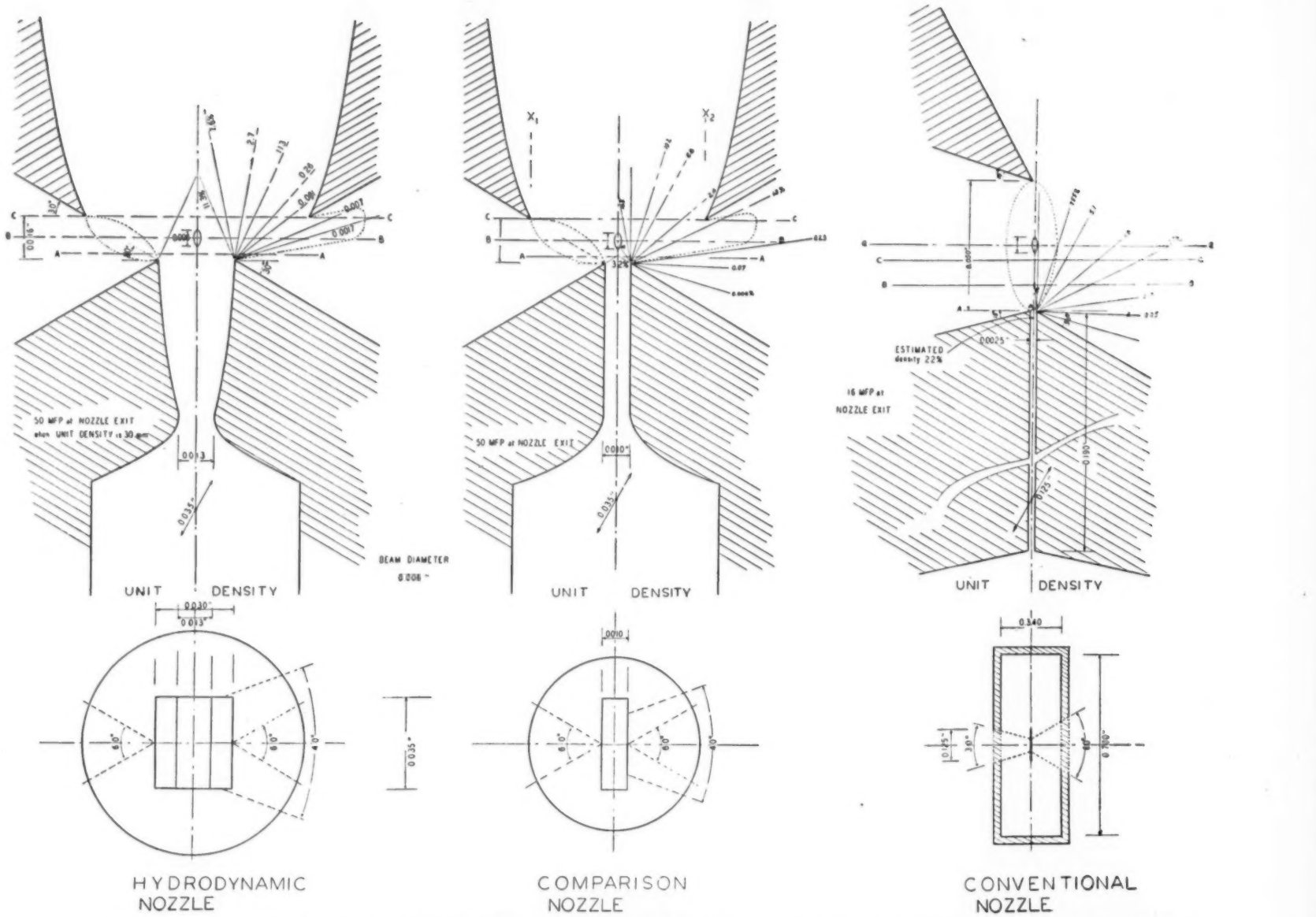


FIG. 2. Vertical and cross sections. Rectangular nozzles designed for two-dimensional flow. The computed isodensity lines just above the exits are given in percentages of the density present in the reservoir. The circle 0.006" diameter shows the position and relative dimension of the electron beam. The electrons proceed from left to right in the figures. The scale for the cross section of 2c is smaller than that of the other figures.

whereas, if liquid air cooling is used, process C_{2b} becomes negligible, and

$$E_{ab}{}^{la}=\frac{C_1+C_3+C_4}{C_1+C_{2a}+C_3+C_4}.$$

The beam absorption data mentioned above are probably more valuable as a means of obtaining an estimate of the extent of multiple scattering relative to the total scattered intensity, once the theoretical relation between these quantities has been established.

From the above discussion it is apparent that an estimate of the merit of a nozzle with respect to sample distribution alone or to multiple scattering alone can be obtained for any specified set of operating conditions as functions of time.

Before these measurements are discussed in detail, it is well to point out that a third definition of efficiency as a function of s, can be obtained from photographic data. This is

$$E_s=\frac{(C_1)_r+(C_3)_r+(C_4)_r}{(C_1)_r+(C_2)_r+(C_3)_r+(C_4)_r},$$

where $(C_n)_r$ is the amount of C_n, integrated over the time of exposure, appearing per unit area on the photographic plate at that radial distance r which corresponds to s for an ideal sample.

DESCRIPTION OF NOZZLES

The manner in which gas flows into the electron diffraction camera depends on the design of the nozzle. In these experiments the three nozzles illustrated in Fig. 2 were tested. First is the so-called "hydrodynamic" nozzle designed by Dr. A. Kantrowitz, of the Graduate School of Aeronautical Engineering of Cornell, to have supersonic flow; the second is referred to as the "comparison" nozzle and has throat dimensions similar to those of the hydrodynamic nozzle but a parallel walled flow channel; the third is the "conventional" nozzle which also has a parallel walled flow channel of about the same cross-sectional area as that of the comparison nozzle but with ratio of width perpendicular to the electron beam to length parallel to the electron beam equal to about 50.

In Fig. 3 is shown the computed gas density destribution across the mouth of the nozzles at three distances from the outlet, A–A, B–B, and C–C. For the conventional nozzle a further line was computed along Q–Q. These graphs show the degree of spreading of the gas from the nozzle mouth, and it is clear that the hydrodynamic nozzle tends to give the best directed flow of gas while the conventional nozzle acts more as a source of gas effusion. In some experiments to be described later a *constriction* was placed in the entrance to the *comparison* nozzle, so that the rate of gas flow was reduced and was more along the order of the conventional nozzle.

It is reasonable to expect that little gas should escape from the side ports (through which the electron beam passes) of the hydrodynamic nozzle, but more from the comparison nozzle, and a relatively large amount from the conventional nozzle. This is really a question of the components of molecular velocity perpendicular to the direction of the electron flow. By measuring the flow rates for the three nozzles under similar conditions, and comparing the observed rates with those calculated from flow formulas, we obtained a rough estimate of the pressure drop along the nozzle. Comparison of the estimated pressure drop with that obtained from the inlet and outlet pressures used in the experiments, gave a measure of the number of molecular collisions on the side walls, as the gas proceeds along the nozzle and thus an estimate of the components of velocity perpendicular to the nozzle axis.

It was shown that the hydrodynamic nozzle had virtually no pressure drop along its length; and only about 5 percent of the applied pressure differential

FIG. 2. Vertical and cross sections. Rectangular nozzles designed for two-dimensional flow. The computed isodensity lines just above the exits are given in percentages of the density present in the reservoir. The circle 0.006" diameter shows the position and relative dimension of the electron beam. The electrons proceed from left to right in the figures. The scale for the cross section of 2c is smaller than that of the other figures.

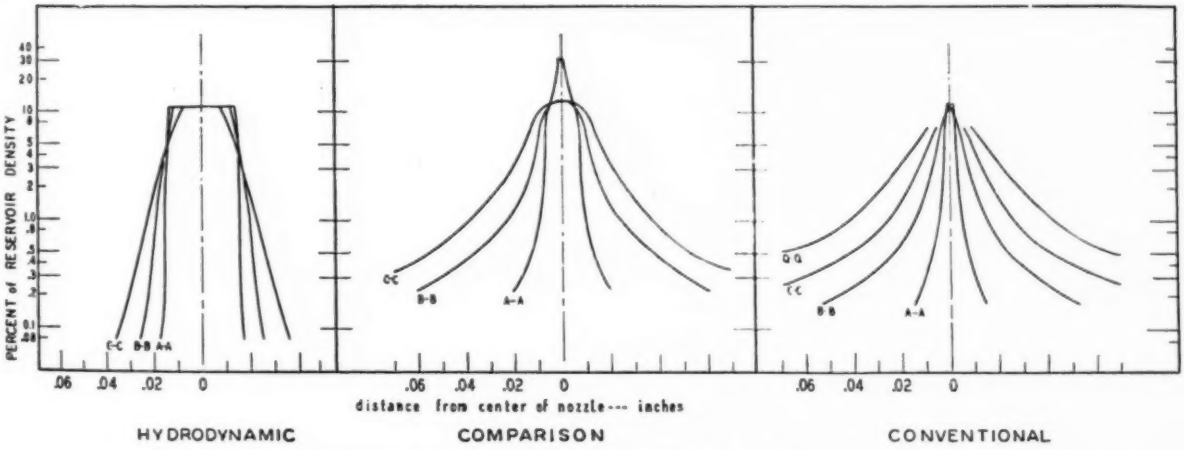


FIG. 3. Computed densities along the ray lines A–A, etc. The values for Fig. 2a are reliable; those for 2b are approximate; those for 2c are little better than guesses.

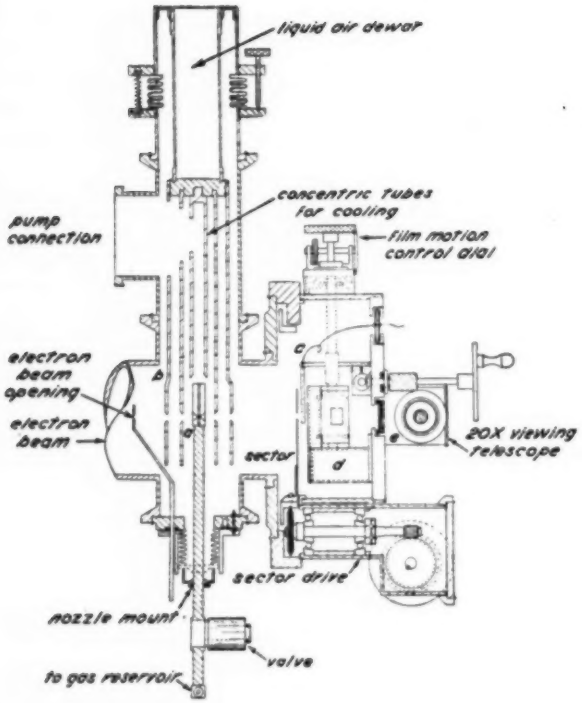


FIG. 4. Side view of electron diffraction apparatus, showing nozzle mount (a), concentric tubes for cooling and condensing the sample (b), pump connection, roll film camera, and sector drive. (c) is a Faraday cage connected to a d.c. amplifier for electron beam control; (d) is a tube carrying a fluorescent screen which is pushed down as the film frame is set in place—the screen remains coplanar to the film; (e) is a short range telescope for viewing the focused spot when the fluorescent screen is in the central position.

was taken up along the walls in the case of the comparison nozzle. Hence the gas exits from the nozzle mouth at a density close to that present at the inlet, and most of the pressure differential is dispersed in the region beyond the nozzle. For the conventional nozzle, one-third to one-half of the applied pressure differential was taken up along the nozzle channel and so the gas must leave the nozzle with comparatively large velocity components perpendicular to the direction of flow. The question may be raised as to what fraction of the gas stream intercepts the electron beam as it fans out from the mouth of the nozzle. It can be shown that under usual operating conditions, most of the gas passes through the beam although with the conventional nozzle the molecules on the average remain in the beam longer so that if comparable gas densities in the diffraction region and exposure times were used for the three nozzles, the observed photographic density would be highest with the conventional nozzle but so would the relative extent of multiple scattering and sample diffusion.

Figure 4 shows a cross-sectional view of the nozzle as placed in the camera. Three concentric chimneys which surround the nozzle can be cooled by filling the metal dewar with liquid air. They also serve to lead the injected gas directly away from the nozzle to the diffusion pump.

THE CALCULATION OF Ep AND THE EFFECTS OF SAMPLE DIFFUSION ON RESOLUTION

The equation for the rate of pressure decrement due to flow of gas at pressure P_s from a volume V_s through a nozzle, into a volume V_c at pressure P_c, may be written

$$-\frac{dP_s}{dt}=\frac{kAu}{V_s}(P_s-P_c)=j(P_s-P_c),$$

where A = area of opening in nozzle, u = velocity of sound, k = flow constant for nozzle, $j=kAu/V_s$. If the volume V_c represents an electron diffraction camera being pumped out at a rate proportional to the pressure P_c, we also must have

$$\frac{dP_c}{dt}=\frac{V_s}{V_c}j(P_s-P_c)-lP_c,$$

where l is a constant depending on the speed of the evacuating pumps. Solution of the above system of differential equations leads to the result

$$P_c=2\frac{mP_0}{b}\left[\exp\left(-\frac{m+l+j+b}{2}t\right)-\exp\left(-\frac{m+l+j-b}{2}t\right)\right],$$

where $m=jV_s/V_c$, $b=((m+l+j)^2-4lj)^{\frac{1}{2}}$, and P_0 is the value of P_s at $t=0$. For the case where V_c/V_s and j/l are much larger than one, this expression may be approximated by

$$P_c=(P_0V_s/V_c)(1-e^{-jt})e^{-lt}. \tag{2}$$

This result is the one which is used in the following considerations.

Since the factor $(1-e^{-jt})$ in the above expression represents the fraction of the gas originally in V_s that has passed through the nozzle in time t, the value of Q_{nozzle} in the expression for E_p may be written

$$Q_{\text{nozzle}}=(1-e^{-jt})(2r/W_n)(\pi r/2\mu),$$

where $(\pi r/2\mu)$ is the average time spent by molecules in traversing an electron beam of radius r, when their average component of velocity perpendicular to the beam is μ, while W_n is the width of the nozzle opening.

$$Q_{\text{camera}}=(\pi r^2R/V_c)Z,$$

where R is the electron beam path length through the camera and

$$Z=\int_0^t(1-e^{-jt})e^{-lt}dt.$$

Hence

$$E_p=\frac{(1-e^{-jt})(2r/W_n)(\pi r/2\mu)}{(1-e^{-jt})(2r/W_n)(\pi r/2\mu)+(\pi r^2R/V_e)Z}. \quad (3)$$

Aside from apparatus constants our equations require the evaluation of the exponential constants "j" and "l." j is a function of the flow rate through the nozzle and l is a function of the rate of evacuation of the camera; hence two sets of measurements were required. To determine j a large fixed volume of gas was allowed to expand through the nozzle into a vacuum, and the pressure of the volume was recorded as a function of time. From the formula $ln(P_2/P_1)=-j(t_2-t_1)$ we get $j=[-ln(\frac{1}{2})]/t_{\frac{1}{2}}$, wherein $t_{\frac{1}{2}}$ is the time at which $P_2=\frac{1}{2}P_1$.

The determination of l was a little more complex, and to do this an ionization gauge was connected to the electron diffraction camera. The output of the ion gauge was fed to a d.c. amplifier and a permanent record obtained with a Brush recording oscillograph. The variation of pressure with time which was thus obtained is shown in Fig. 5 for a typical case. The sharp rise is characteristic of the factor $(1-e^{-jt})$ and the exponential decrease is due to the (e^{-lt}) factor. The value of l can be determined from the curve since the camera pressure at any time is given by $P=(P_0V_s/V_c)\times(1-e^{-jt})(e^{-lt})$. The calculation of all the constants in the efficiency equation is now straightforward and efficiency can be calculated as a function of time.

Pressure measurements were made under two sets of conditions. By means of the cooling chimneys (Fig. 4) we can condense out most gases before they can diffuse throughout the camera. The curves in Fig. 5a are for the case of the chimneys at room temperature, while those in Fig. 5b are for the chimneys cooled with liquid air. The pressure scale is compressed by a factor of about 5, and the time scale by a factor of about 10 in the latter. Efficiencies were calculated as a function of time, with and without the use of liquid air cooling, and the results plotted in Fig. 6. The following may be pointed out:

(a) E_p begins at 100 percent for $t=0$, and drops for increasing t.

(b) The E_p of a nozzle without adequate liquid air cooling is less than that for a nozzle with adequately cooled condensing surfaces around it.

(c) The comparison nozzle shows the greatest E_p of the three for no liquid air cooling—probably due to the fact that the molecules remain in the beam a longer time than in the case of the supersonic nozzle.

(d) The conventional nozzle is least efficient up to 0.6 second regardless of conditions.

(e) It would follow from the efficiency equations that a high pumping rate would favor higher efficiencies. Liquid air cooled surfaces fulfill this requirement to a substantial extent for most gases. Similarly a *small* camera volume would be desirable, or else one so large that the injection of a gas sample would not raise the pressure appreciably in the camera.

The appearance of electron diffraction photographs was correlated theoretically to the value of the efficiency E_p by comparison of the integrated background intensity (atomic contribution) and of the molecular contribution for an extended sample with the corresponding quantities expected from an infinitesimally thin sample. It was found that when no beam trap is inserted close to the principal region of diffraction, and the sector is placed at the plane of the detector, little departure from unit background is introduced (<2 percent) if the sample is relatively concentrated, as in (a), (c), and (d), Fig. 7. However, for a uniformly distributed sample (Fig. 7, curve *e*), the density at small s is 15–20 percent higher than at large s. The computed effect on the molecular terms due to diffusion of the sample is quite marked. Inspection of the curves in Fig. 8a, b, c will show that due to the overlapping of patterns arising at different sample-detector distances (process 2, Fig. 1) the high frequency terms are more readily damped at large "s" than are the lower frequency contributions. As a consequence the unwary investigator will be led to assign too large temperature factors to the long inter-atomic distances. The above integrations were carried out graphically.

THE CALCULATION OF E_{ab} AND THE EFFECTS OF SECONDARY DIFFRACTION ON RESOLUTION

An estimate of E_{ab} was obtained from measurements of the loss of current in the main electron beam after

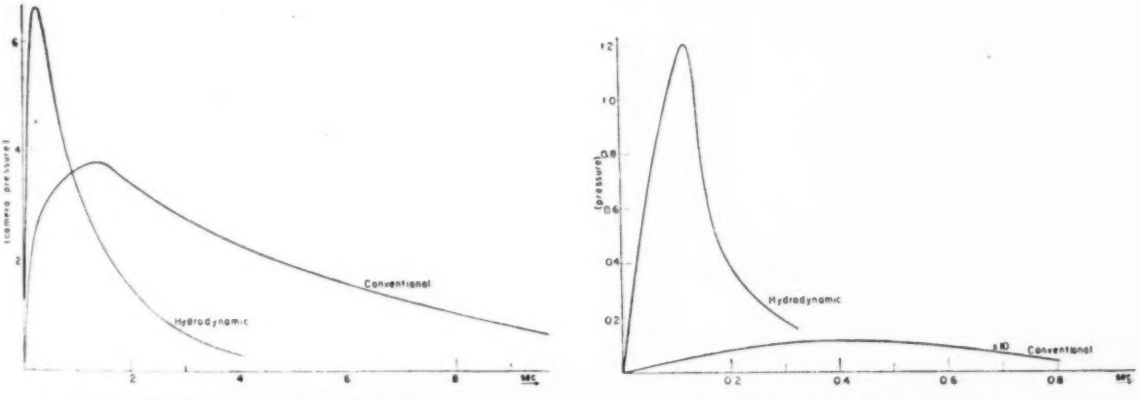


FIG. 5. Pressure transients in the camera. (a) Chimneys at room temperature; (b) Liquid air-cooled chimneys.

it had passed through an injected gas sample under the following set of arrangements:

(a) The main beam was allowed to pass through the nozzle, with the surrounding chimneys at room temperature. Absorption of the beam is by processes 1, 2, 3, and 4.

(b) The nozzle was tilted so that the main beam passed beside it; the nozzle surroundings remained at room temperature. Absorption is by process 2.

(c) The main beam was allowed to pass through the nozzle but the surrounding chimneys were liquid air-cooled. This measures absorption by processes 1, 3, 4, and that part of 2 between nozzle and condensing surfaces (process 2_a).

(d) The main beam was allowed to bypass the nozzle as in (b) but the surrounding surfaces were cooled. Absorption of the beam is by diffraction between the nozzle and the condensing surfaces (process 2_a).

On the basis of measurements such as these one can compute E_{ab} as the ratio of diffraction over the nozzle to the total diffraction at any time during which the sample is in the camera.

The absorption measurements were made with the aid of a d.c. amplifier and Brush recording oscillograph connected to a collecting electrode (cup) of five millimeter diameter placed in the path of the main electron beam at a distance of 19 cm from the nozzle. In Fig. 9b a typical curve of the variation of the main beam current with time is shown for the cases cited above. In curve 1 we have the case of maximum absorption followed by a fairly rapid decay. Case (2) has a much smaller maximum, but it is shifted appreciably to the right indicating, possibly, that gas diffusion around the condensing surfaces and through the camera is a comparatively slow process. Case (3) shows a very sharp rise and rapid decay, for the diffraction taking place over the nozzle. Case (4) failed to give any measurable change in beam current which indicates that absorption of the beam by ambient gas within the cooled condensing surfaces was so small that it was not detected by our instrument. Figure 9a shows a curve of efficiency calculated from the absorption curves.

A fairly good comparison of the nozzles is obtained by quoting $t_{\frac{1}{2}}$ values, i.e., the time for the maximum of the curve to be reduced by 50 percent. See Table I.

In Table II are given the percentages absorption of the beam for the case of three nozzles under various experimental conditions. The values are those of maximum absorption, and the times for maximum values to be reached are also quoted.

The comparison and hydrodynamic nozzles are closely similar and both are more effective on the basis of E_{ab} values than a nozzle of slow gas admittance. If one computes the ratio [diffraction over the nozzle/diffraction elsewhere in the camera] one finds a value of 12 to 13 for the comparison and hydrodynamic nozzles and only about 1.5 for slow gas admittance. With liquid air cooling, the absorption in the camera was not measurable on our apparatus and the E_{ab} values were close to unity in all cases.

It must be pointed out that slower gas flow rates mean lesser absorption of the main beam and as a result less multiple scattering.

In order to use the absorption measurements as a means of estimating the extent of multiple scattering, the following relation, derived in Appendix C, may be used:

$$J_2/J_1=1+[(J_0/J_1)-1]\ln[1-(J_1/J_0)], \qquad (4)$$

where J_0 is the total intensity of the incident main beam, J_1 is the total intensity of scattered radiation, and J_2 is the total intensity of electrons which have been scattered more than once. For values of J_1/J_0 up to about 0.2, J_2/J_1 is fairly well approximated by 0.5 J_1/J_0; for J_1/J_0 equal to 0.3 and 0.5, J_2/J_1 equals 0.56 J_1/J_0 and 0.61 J_1/J_0, respectively. Since J_1/J_0 is measured directly in the absorption experiments, we

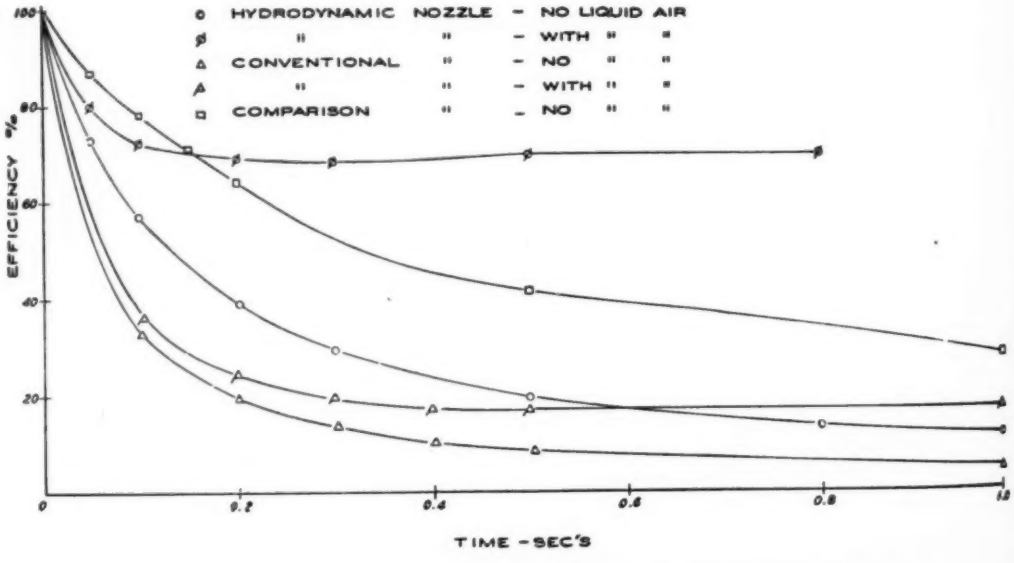


FIG. 6. Nozzle efficiencies (E_p) as a function of the time.

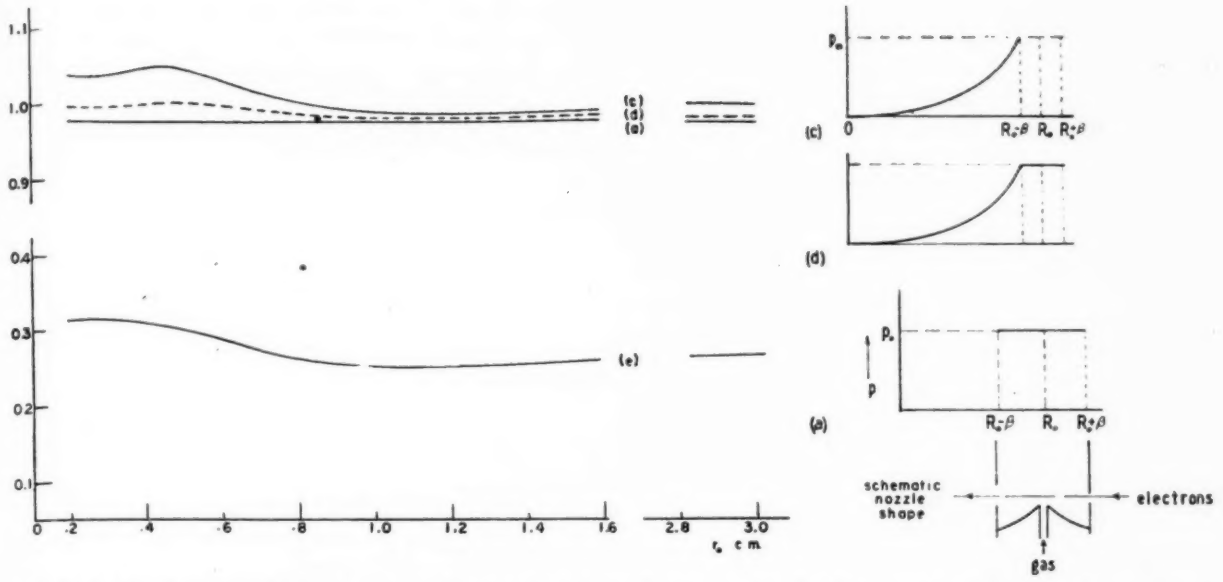


FIG. 7. Relative background intensity curves, for an extended sample compared to that for a sample of the same size concentrated at R_0. The sector is at the plane of the detector, and no beam trap is interposed. The curves correspond to the pressure distributions indicated below, with $R_0=19.33$ cm. $\beta=0.475$ cm. $p=[(p_0\beta^2)/(R_0-R)^2]$ for $R\leqslant(R_0-\beta)$. r_0 is distance along the plate. Curve (e) is for a uniform pressure distribution throughout the camera.

have a very simple way of estimating the relative extent of multiple scattering.

For rapid gas admittance (hydrodynamic nozzle), as much as 30 percent of the incident main beam may be absorbed. This means that J_1/J_0 equals 0.30 and hence that the ratio of the amount of multiple scattering to total scattering equals 0.56 times 0.30, or 0.17. For the case of slow gas admittance where the maximum beam absorption may be only 2 percent; we find then that J_2/J_1 equals 0.50 times J_1/J_0, i.e., 0.01, so that the relative extent of multiple scattering has been reduced by a factor of 17 as compared with the case of rapid gas admittance.

In order to correlate these results with the observed appearance of electron diffraction photographs, we must estimate the ratio of the loss of intensity, (I^*-I), resulting from multiple scattering in a ray reaching a particular point on the plane of the detector, to the gain, I_n, of background intensity resulting from the contribution at this point of intensity rescattered from all other rays. In formulating the integrals for this additional background it soon became clear that the general case is very involved; hence we have simplified the discussion by not considering the dependence of the result on the degree of sample extension in the direction of the electron beam. Thus it is assumed that the sample is restricted to a region $(R_0+\beta)$ to $(R_0-\beta)$ directly over the nozzle, and that $\beta/R_0\ll1$ so that rays diffracted at an angle θ originating in any part of the sample all arrive on the plane of the detector at essentially one distance from the central beam. It was also assumed that the gas density is sufficiently low that all of the multiple scattering may be considered to be secondary scattering and that the intensity of the main beam may be considered constant while passing through the sample. Finally, it was found convenient to approximate the variable "s" by $2\pi r/(\lambda R_0)$ where $r/R_0=\tan\theta$. The particular molecule used as an example was carbon tetrachloride, and for it the following empirical relation was fitted:

$$\text{(Atomic contribution + molecular contribution)}\times(2\pi/\lambda R_0)^4 1/s^4=\frac{4280}{(s^2+22)^2}+\frac{310}{(s^2+0.6)^2}. \quad (5)$$

The values of $I_n/(I^*-I)$ which result from these calculations are plotted in Fig. 10. It is apparent that the background intensity is decreased at small s and increased at large s as a result of secondary scattering. It may be demonstrated that the *fractional loss* is independent of s (Appendix A).

To consider a definite example, suppose that the measured absorption of the beam is 20 percent, so that the ratio of the amount of multiple scattering to total scattering is 0.1. Then, assuming for simplicity that at some value of s, say $s=8$, the background intensity I^* that would be obtained if there were no multiple scattering is equal to one, the actual contribution, I, from electrons scattered only once will be $(1.0-0.1)=0.9$, while the contribution I_n resulting from multiple scattering of all other rays will be roughly $2(I^*-I)=0.2$. Hence the total background intensity will be $0.9+0.2=1.1$. Therefore if the ratio [(molecular contribution)/(atomic contribution)] is equal to x under conditions where no multiple scattering occurred, it

will only be equal to $(0.9/1.1)x$ under the actual conditions stated. Indeed, the value will be even less than this because we have not taken into account the fact that rescattering of the molecular contribution will result in a preferential addition of intensity to the valleys of the molecular contribution curve.

Because the background contribution to the total intensity is large compared to the molecular contribution for large values of s even in the ideal case, it is obvious that every effort should be made to reduce the extension, it is apparent that whether the observed background happens to be too high, flat, or too low for small s depends on the balancing of two opposing factors; extension of the sample tends to raise the background (Fig. 7, e) while secondary scattering tends to lower it (Fig. 10). It is quite clear why some observers who failed to consider these complications in nozzle design were misled by the extraneous backgrounds into believing that the atom form factors are necessarily incorrect.

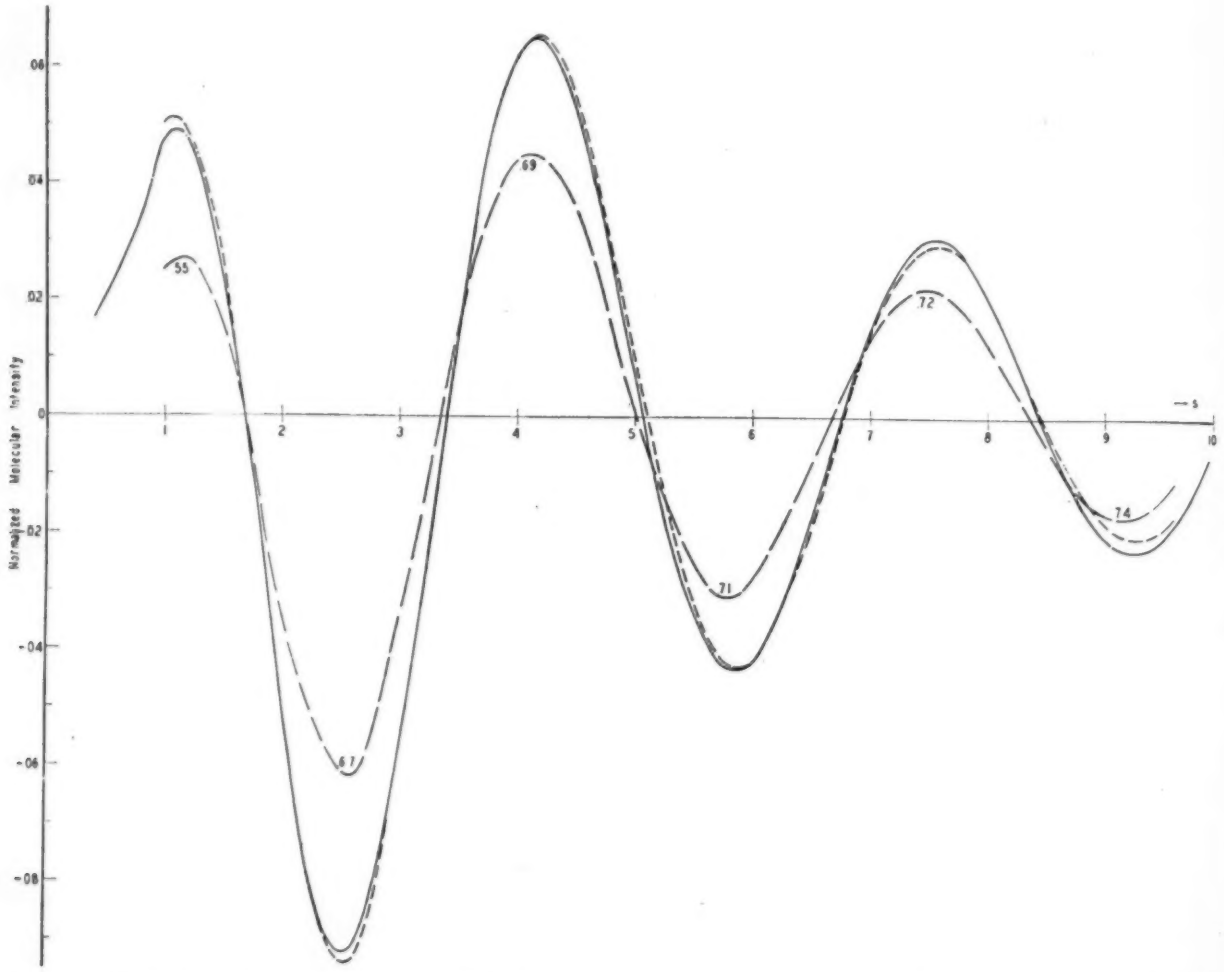


FIG. 8a. Molecular intensity curves for an extended sample using the B–B term of B_2H_6, utilizing a sector and beam trap. Curve Inf. is for an infinitesimally thin sample at $R_0=19.33$ cm; $\lambda=0.06359A°$. $2f_B{}^2e^{-.00418s^2}\times(\sin 1.86s)/1.86s$. In curve a we assumed that the pressure distribution was uniform, extending from $(R_0+\beta)$ to $(R_0-\beta)$, and zero everywhere else. $\beta=0.475$ cm. In curve b the pressure distribution was assumed to be uniform, and to extend from R_0 to $(R_0-\beta)$; i.e., an unsymmetrical nozzle. Note the effect of the beam stop on the positions of the maxima and minima, and on their relative intensities. The numbers on this curve indicate the fractional intensities of the corresponding peaks compared to curve Inf. [The broken line curve should be designated a above the abscissa and b, below. The solid line curve should be marked Inf.]

beam absorption in the nozzle to a low value. To do this requires of course that the sample density be decreased and that the exposure times be correspondingly lengthened to an extent consistent with the limitations imposed by stability of the electron gun, high voltage, pumping rates, etc.

Also, from these results and the ones for finite sample

PHOTOGRAPHIC EVIDENCE OF LOSS OF RESOLUTION

A series of photographic records (i.e., actual electron diffraction pictures) was made using cooled condensing surfaces in some exposures and none in others, and tilting the nozzle so that diffraction by gas in the camera could be measured. We can thereby define an

efficiency, E_s. If the photographic plate is read with a microphotometer and a record obtained, a line can be drawn through the fluctuations of the curve in such a way as to represent the background. Comparison of this background when the beam passes through the nozzle (case of normal diffraction procedure) to the background when the nozzle is tilted and the beam passes through the gas in the camera only, gives

$$E_s=\frac{D_{\text{nozzle}}}{D_{\text{nozzle}}+D_{\text{camera}}}=\frac{(C_1)_r+(C_3)_r+(C_4)_r}{(C_1)_r+(C_2)_r+(C_3)_r+(C_4)_r}.$$

E_s values can be determined for the case of liquid air cooled surfaces around the nozzle and for the surfaces at room temperature. The measurements were made on Eastman Commercial plates when no sector was used, and on Eastman Microfile when a sector was used. Sensitometric curves were determined on both of these emulsions so that the photographic densities could be converted to electron intensities. The photographic densities were measured with a Leeds and Northrup recording microphotometer. CCl_4 vapor was used as a condensable gas throughout the experiments.

E_s values are plotted in Fig. 11 for the hydrodynamic nozzle and for the comparison nozzle with a constriction so that the gas entered slowly. The hydrodynamic nozzle with the surrounding chimneys at room temperature allows a considerable amount of gas to diffuse through the camera so that the efficiency is considerably less than 100 percent. The observed dropping off of the efficiency with decreasing s should be expected because of the fact that the background contributed by that portion of the sample which has diffused through the camera falls off more rapidly with s than that contributed by an ideal sample, and because of the fact that multiple scattering in the dense gas directly over the nozzle lowers the background contributed by it at small s values. Therefore from the standpoint of this latter factor, the apparent increase in efficiency with s is actually an indication that resolution is *reduced* at large s values. With liquid air cooling the efficiency remains at about 99 percent irrespective of the angle of diffraction; this indicates that the liquid air-cooled shields effectively remove the diffracted gas but implies nothing about the seriousness of multiple scattering. It is interesting to note that the nozzle of slow gas admittance (comparison nozzle plus constriction) gives better E_s values than the hydrodynamic nozzle. This may be a result of more efficient operation of the evacuation pumps when the pressure in the camera is kept low, as will be the case if the gas is injected into the camera slowly. With the chimneys liquid air-cooled, efficiencies of about 99 percent were again obtained for slow gas admittance.

The actual resolution for the diffraction patterns can be determined by reading the height of the peaks above the valleys in ratio to the background and comparing this to the theoretical ratio of peak height to background. These figures give us, then, the loss on intensity of the peaks and gain of intensity of the background. That secondary scattering and diffraction by an extended sample do lead to a loss in resolution can be demonstrated in the following simple way. Suppose

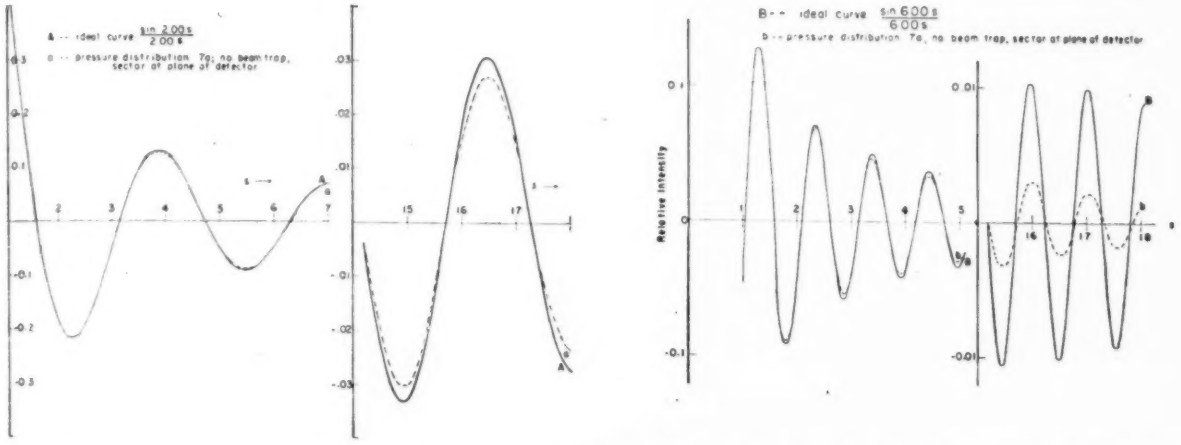
$$I_{\text{theoretical}}(s)=A_1(s)+A_2\sin\gamma s$$

with $A_2<A_1(s)$, which is a slowly varying function. Let the fraction $(1-\beta)$ be lost through multiple scattering, and let the total gain in background be $[B_1(s)+B_2(s)]$. Then

$$I_{\text{observed}}(s)=[B_1(s)+B_2(s)]+\beta[A_1(s)+A_2\sin\gamma s].$$

A measure of the resolution is the magnitude of the observed "relative fluctuation," i.e.,

$$\frac{2\beta A_2}{[B_1(s)+B_2(s)]+\beta A_1(s)}\left[\frac{2A_2}{A_1(s)}\right]^{-1}\leqq 1-\frac{B_1(s)+B_2(s)}{\beta A_1(s)}.$$



FIGS. 8b and 8c. No beam trap correction was inserted in these integrations, and the sector was placed in the plane of the detector. The pressure distributions [they should be marked 8*a* rather than 7*a* in both parts of the figure] postulated were the same as for Fig. 8a, curve *a*.

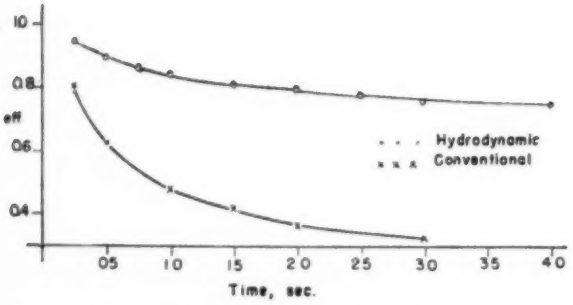


FIG. 9a. Nozzle efficiencies (E_{ab}) as a function of the time.

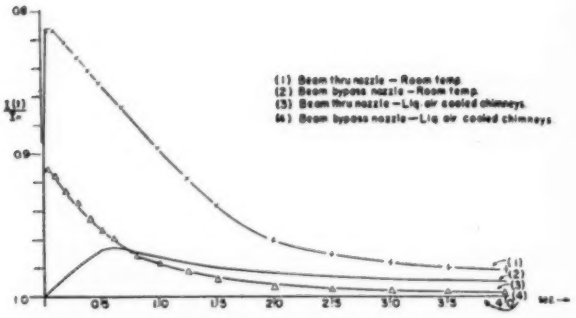

(1) Beam thru nozzle – Room temp.
(2) Beam bypass nozzle – Room temp.
(3) Beam thru nozzle – Liq. air cooled chimneys.
(4) Beam bypass nozzle – Liq. air cooled chimneys.

FIG. 9b. Records of beam current absorption curves.

Thus, the resolution is diminished by a factor which is inversely proportional to the fraction of each ray which gets by without suffering secondary scattering, and directly proportional to the gain in background. Further, it is probable that $B_2(s)$, which arises from the secondary scattering of the A_2 term, is a harmonic function of s, but with a phase which does not match that of $\sin\gamma s$; this effect will introduce a further decrease in resolution.

By measuring the beam absorption and computing the ratio of peak height to total intensity (at the 7th peak for CCl_4) for a series of inlet pressures we obtained the results listed in Table III. This table shows very clearly the effect of the rate of gas admittance on the resolution since we reduced our efficiency by a factor of 3 for increase of gas pressure in the reservoir by a factor of 7 or so. As the beam absorption increases from 8 percent to 16.7 percent, the resolution, as measured, drops by a factor of almost 2. Actually, our beam absorption figures are smaller than they should be, for the Faraday cage introduced to remove the main beam between the nozzle and the detector must be larger than the beam in order to avoid edge scattering. Hence a part of the diffracted pattern is trapped with the main beams when gas is admitted.

To conclude this section it would be in order to say a word about the effects on resolution of imperfect focus at the plane of the detector. A large scale microphotometer record was prepared from an etched gold foil diffraction pattern. From this the distribution of intensity over a ring, $K(s_0-s)$, was determined, due not only to the finite size of the beam but also to the finite size of the slit in the L and N microphotometer. The observed net effective width at half maximum for the (111) ring was 0.03 cm,[4] with $R_0=19.33$ cm. By comparing the integrated $\int K(s_0-s)I(s)ds$, with the ideal $I(s)$ gas diffraction patterns, it was found that:

(1) there is no lateral shifting of the maxima and minima;
(2) the ratio of the observed peak heights to the theoretical peak heights is less than unity, i.e., there is a loss of resolution;
(3) This ratio is independent of s, but decreases with increasing interatomic distances. This effect, similar to that arising from extension of the sample, also leads to the assignment of false temperature factors.

Interatomic distances	1.86	5.58	7.44
Ratio $\frac{\text{integrated}}{\text{ideal}}$	0.97	0.84	0.72.

Actually, a slight dependence of resolution on s is introduced in most electron diffraction apparatus through the use of a flat plate instead of one curved to the focal circle. However, the additional broadening is quite small (at $s=23$ the effective increase in half-width of the diffraction peak is less than 1 percent).

TABLE I. Values of $t_{\frac{1}{2}}$ for beam absorption measurements.

Nozzle	No liquid air		Liquid air	
	In beam	Out of beam	In beam	Out of beam
Hydrodynamic	0.5 sec.	1.3 sec.	0.36 sec.	0.00 sec.
Comparison	0.57 sec.	2 sec.	0.45 sec.	0.00 sec.
Comparison +constriction	6 sec.	6 sec.	2 sec.	0.00 sec.

A MOLECULAR BEAM NOZZLE

From the data presented above, one is led to speculate on the desirability of using a molecular beam in order to avoid the effects of secondary scattering and spreading of the gas through the camera. The principal difficulty in using a conventional molecular beam is that the gas densities are not sufficiently high to avoid making unmanageably long exposures. However, it is

TABLE II. Absorption of main electron beam and time to maximum absorption.

Nozzle	No liquid air		Liquid air	
	In beam	Out of beam	In beam	Out of beam
Hydrodynamic	31 % at 0.08 sec.	2.1% at 0.37 sec.	27 % at 0.06 sec.	0.0%
Comparison	35 % at 0.10 sec.	2 % at 0.5 sec.	25 % at 0.10 sec.	0.0%
Comparison +constriction	2.4% at 0.80 sec.	1.1% at 1.5 sec.	1.6% at 0.1 sec.	0.0%

[4] Later, measurement of a trace of the direct beam showed that at half maximum, it is 0.0035 cm wide, so that the effects computed below are much larger than those actually present. The diffraction broadening due to our gold foil is approximately 10 times the diameter of the focused spot.

not necessary that a molecular beam used for electron diffraction work have the degree of collimation required for other types of studies. A possible design for a nozzle satisfactory for electron diffraction is indicated below. Experimentally we found that in order to obtain sufficient photographic density in the diffraction photographs ($D=0.2$ at $s=25$ for microfile film), there must be approximately $(5\times10^{+14})/ti$ molecules in the electron beam per square cm of beam cross-sectional area, where t is the duration of the exposure in seconds, and i is the total beam current in microamperes.

Let us begin by supposing that a source slit S_1 (see Fig. 12) of dimensions 0.01×0.1 cm is used. The small dimension of such a slit determines the maximum gas sample density which may be used without obtaining an appreciable number of collisions in the gas stream beyond S_1. The maximum density is the one at which the gas molecules have a mean free path equal approximately to the small dimension of the slit. Thus for a slit 0.01 cm wide, the maximum sample pressure which one may safely use is about 0.5 mm; or N, the number of molecules per cc in the reservoir, is about 1.5×10^{16}.

The gas density which appears directly above S_2, where the electron beam passes by, is approximately[5] $N'=Na/4\pi l^2$ where N' is the gas density in molecules per cc at S_2, a is the area of the source slit S_1, and l is the separation of S_1 and S_2. However, what is actually required is not N', but N'', the number of molecules in the electron beam per square cm of beam cross section; this quantity may be obtained simply by multiplying N' by that dimension of S_2 which lies along the direction of the electron beam. Therefore if this dimension is 0.1 cm, we find that

$$N''=\frac{1.5\times10^{16}\times1.0\times10^{-3}}{12.5l^2}\times0.1=\frac{1\times10^{11}}{l^2}.$$

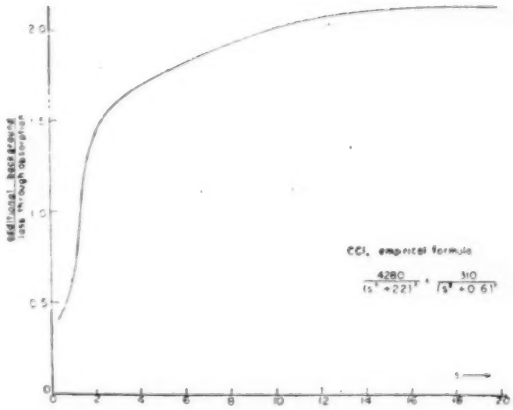


FIG. 10. Computed values for the ratio

$$\frac{\text{Addition background due to secondary scattering}}{\text{Loss in background due to absorption}}$$

as a function of s.

[5] R. G. J. Fraser, *Molecular Rays* (Mcamillan Company, Inc. New York, 1931), p. 14.

TABLE III. Observed resolution.

CCl_4 press	Absorption % (as measured using 5 mm collector)	Peak height / Total intensity ×100	Resolution (%)
—	0	12.0 (theoretical)*	100
16.5 (−12°C)	8	11.5	96
32.9 (0°C)	16.7	6.8	57
119.5 (26°C)	46.5	3.9	33

* Note: the theoretical curve was constructed using the temperature factors $e^{-0.0016s^2}$ and $e^{-0.0021s^2}$ for C−Cl distances and Cl−Cl distances, respectively.

From this relation the approximate exposure times required to give satisfactory electron diffraction photographs can be calculated as a function of l. Assuming that the main beam current i equals one microampere, we thus find that the time of exposure t equals 2×10^4, 800, and 110 seconds for l equal to 2, 0.4, and 0.15 cm, respectively.

From the standpoint of short exposure times one would like l to be small. On the other hand, if l is made very small, not only will the angular divergence of the beam be large, but also molecules which hit the bottom edges of S_2 will require too many additional bounces from the surfaces of S_1 and S_2 to reach the wide throat of the pump; the greater the average number of bounces required, the greater will be the probability for any one of these molecules to go through S_2 along an undesirable path rather than to the diffusion pump. For this reason we wish to select a compromise value—to make S_2 wide enough to allow a good fraction of the molecules leaving S_1 to go directly through S_2, but still to maintain molecular beam characteristics. The practical limit is determined by the maximum sample extension which is to be allowed. With $l=0.15$ cm and S_2 0.1×0.1 cm, the total angular divergence will be about 42° so that the maximum sample extension along the direction of the electron beam at a distance 0.1 cm above the bottom of S_2 will be 0.2 cm, a not unreasonable value. Roughly 55 percent of the molecules leaving S_1 will go directly through S_2 in this case.

Table IV summarizes some of the characteristics and requirements of this type of molecular beam nozzle, and Fig. 12 illustrates its physical outlines.

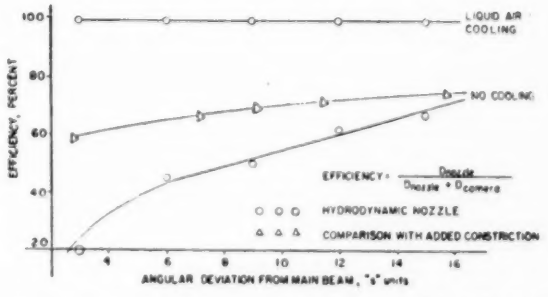


FIG. 11. Nozzle efficiencies from photographic data (E_s), as a function of s.

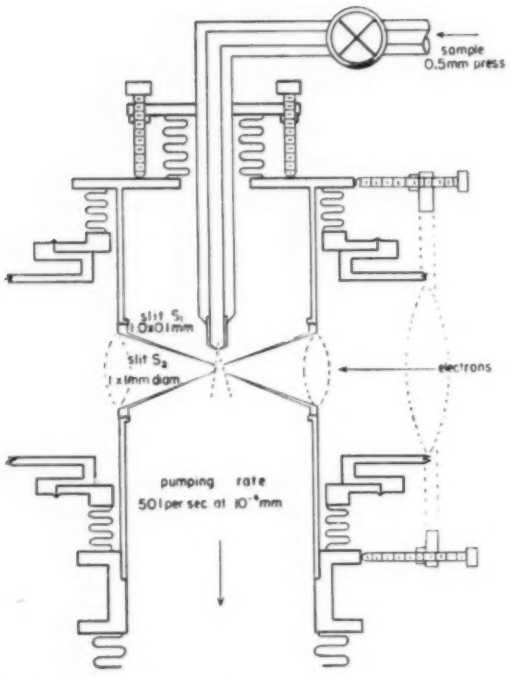


FIG. 12. Proposed molecular beam nozzle and pumping section.

If the gas densities achieved by this nozzle are found to be too low, or the angular divergence too great, a source of high gas density has been designed by Dr. A. Kantrowitz[6] of the Graduate School of Aeronautical Engineering at Cornell, utilizing supersonic flow. Under comparable conditions a molecular beam intensity about 75 times that obtainable with standard molecular beam geometry can be produced. However, due to its added complexity and inefficient sample utilization it would not be advantageous unless the factors of higher gas density, and lower effective temperature of the gas (approximately 72°K for a Mach number 4) were necessary for the investigation in hand.

TABLE IV. Properties of modified molecular beam nozzle.*

Property	Calculated value	
Dimension of slit S_1	0.01 × 0.1 cm	
Dimension of slit S_2	0.1 × 0.1 cm	
Pump capacity (on nozzle—Fig. 12)	50 l/sec. at 1×10^{-4} mm	
Molecules in electron beam per unit area of electron beam	6×10^{11}	
Molecular velocity	5×10^{4} cm per sec.	
Effective molecular temperature	300°K	
Consumption of sample	3×10^{-7} moles/sec.	
Exposure time and total beam divergence	time	angle
$S_1 - S_2$ distance 2 cm	5.5 hrs.	3°
4 mm	13 min.	14°
1.5 mm	2 min.	42°

* With beam current increased by a factor of 5 (to 5 microamperes) the exposure time in the last case will be reduced to 24 seconds.

[6] J. Grey and A. Kantrowitz, Phys. Rev. **75**, 1331 (1949). For details refer to Master's Dissertation submitted by Mr. Grey to the Graduate School of Aeronautical Engineering at Cornell University (1949).

DISCUSSION AND CONCLUSIONS

We have attributed the lack of resolution of the diffracted electron image on the detector to failure of assumptions (2), (3), and (4) listed in the introduction. The data obtained may be summarized as follows:

Pressure Measurements

(i) $E_p(l)$

$$= \frac{\text{molecules coming from the nozzle}}{\text{molecules from nozzle} + \text{molecules from camera}}$$

(ii) Better efficiencies are given by the hydrodynamic or comparison nozzles than the conventional, with or without the use of liquid air.

(iii) Efficiencies are always higher for shorter exposures, and for high camera pumping speeds.

(iv) It is desirable to beam the gas through the camera and directly into a diffusion pump.

Beam Absorption Measurements

(i) $$E_{ab}(l) = \frac{\text{beam absorption over the nozzle}}{\text{beam abs. over nozzle} + \text{beam abs. in camera}}$$

This is greater for the hydrodynamic or comparison nozzles than for the nozzle of slow admittance by a factor of 50 percent when no liquid air cooling is used.

(ii) Beam absorption with the hydrodynamic or comparison nozzles may be as great as 30 percent, and secondary scattering within the nozzle will be serious.

(iii) Liquid air-cooled surfaces *do not* reduce secondary scattering over the nozzle exit.

Photographic Measurements

(i) The efficiency is an integrated value over the exposure time, being defined:

$$E_s = \frac{\text{Photographic Density from gas out of nozzle}}{\text{P.D. out of nozzle} + \text{P.D. with beam by-passing nozzle}}.$$

(ii) At room temperature the efficiencies become strongly dependent on the angle of diffraction.

(iii) With liquid air-cooled surfaces around the nozzle the efficiencies are close to 100 percent.

(iv) Secondary scattering may be serious; it is dependent on scattering angle; and the shape of the background can be appreciably changed with density of the sample gas (probably due to the differences in secondary scattering). When absorption of the main beam exceeds about 10 percent, resolution is considerably impaired.

In the analysis of secondary scattering it was shown that an angular dependence existed, although secondary scattering was not dependent on the sample distribution. The ratio of [gain in background/loss of primary diffraction] rises to 1.6 (at about $s=3$), and continues

to increase to a value about 2 for larger values of s. The secondary scattering is chiefly a function of N, the number of molecules available to the beam. The conventional nozzle, because of its slow gas admittance, will give rise to the least multiple scattering but allows a great deal of sample extension to occur. The hydrodynamic nozzle, under optimum conditions, gives a much better defined gas stream but for hydrodynamic principles to apply, a gas density so high that multiple scattering becomes very serious must be used and hence this nozzle has little if any advantage over the comparison nozzle, especially if liquid air-cooled chimneys which reduce the seriousness of sample extension can be used. As a means of avoiding to a large extent the effects of both sample extension and multiple scattering, a design for a modified molecular beam nozzle suitable for electron diffraction work is offered. With any type of nozzle, the escaping gas should be able to flow in a straight path directly into a diffusion pump of high speed.

With proper nozzle and camera design it should be possible to achieve nearly maximum resolution in the electron diffraction patterns of gases.

ACKNOWLEDGMENT

Several grants made the completion of this investigation possible. We wish to thank sincerely: The Canadian Defence Research Board—a fellowship for R. B. Harvey; The Office of Naval Research—a research assistantship for F. A. Keidel (under contract No. N6-ori-213, Task I NR 052-040); The Guggenheim Memorial Foundation—a fellowship for S. H. Bauer. All of us are indebted to Professor A. Kantrowitz for his assistance with nozzle designs.

APPENDIX A. ANGULAR INDEPENDENCE OF ABSORPTION OF DIFFRACTED RAYS

The statement that the absorption of the diffracted rays is independent of s may be justified on the assumption that the actual gas distribution in the vicinity of the nozzle is cylindrically symmetric about the axis along which the main stream of gas flows into the camera, while the incident electron beam enters along a diameter of this cylinder. Then, if J_0 is the intensity reaching a primary beam collector when no sample is present, and $N(R, t)$ is the distribution in the number of molecules per unit volume, as a function of distance from the gas stream axis and time, the primary beam intensity, observed when a sample is present, will be

$$J(t)=J_0\exp\left[-\pi\sigma^2\int_0^{R_0+\xi}N(R,t)dR\right], \qquad \text{(A1)}$$

where σ is the electron-molecular collision radius, and is in a sense defined by the above absorption equation, and $(R_0+\xi)$ is the maximum extension of the sample, so that the limits 0 to $(R_0+\xi)$ include all the gas molecules which intercept the beam, from the last aperture in the magnetic lens to the camera.

If there were no higher order scattering, the intensity per unit area of collector at some point r in the plane of the observer, measured from the undeflected ray, would be

$$I^*(r/R_0,t)=J_0B^2\int_0^{R_0+\xi}N(R,t)\frac{(\text{atomic}+\text{molecular})}{R^2[s(r/R)]^4}dR, \qquad \text{(A2)}$$

where

$$B^2=\left[\frac{8\pi^2me^2}{h^2}\right]^2;\quad s(r/R)=\frac{4\pi}{\lambda}\sin\frac{\theta}{2};\quad \theta=\arctan\frac{r}{R}.$$

Equations (1) and (A2) are identical except that the former is valid only if all the N molecules are located at R_0, whereas in (A2) we have permitted the sample to be distributed as per $N(R, t)$. Thus it gives the intensity for all rays diffracted by an extended sample assuming no secondary scattering. However, higher order scattering introduces an effective absorption so that the actual intensity of electrons scattered only once will be

$$I(r/R_0,t)=B^2\int_0^{R_0+\xi}J_0\exp\left[-\pi\sigma^2\int_R^{R+\xi}N(R,t)dR\right]N(R,t)\times\frac{(\text{atomic}+\text{molecular})}{R^2s^4}\exp\left[-\pi\sigma^2\int_0^{R}N(R,t)dR\right]dR. \qquad \text{(A3)}$$

The first exponential factor allows for the attenuation of the primary ray in proceeding from $(R_0+\xi)$ to R; the second allows for the attenuation of the diffracted ray in proceeding from R to the plane of the detector (at 0). Hence, subject to the assumption that *the angles of diffraction are not large*,

$$I(r/R_0,t)=I^*\exp\left[-\pi\sigma^2\int_0^{R+\xi}N(R,t)dR\right]. \qquad \text{(A4)}$$

Thus *no selective loss* is to be expected for any ray, be it primary or diffracted. Similarly, both the atomic and molecular contributions will be reduced by the same fraction.

APPENDIX B. COMPUTATION OF COLLISION CROSS SECTION

The absorption coefficient $\pi\sigma^2$ may be estimated in the following way: Circumscribe around a sample located in its entirety at R_0, a sphere of radius R_0. Then an element of area on that sphere is given in terms of the solid angle $\Delta\theta\Delta\phi$ by $[R_0\sin\theta\Delta\phi R_0\Delta\theta]$. The intensity of electrons scattered into this element of solid angle by $N_0(t)$ molecules (per unit area of the incident beam),

$$j(\theta,t)=J_0B^2N_0(t)\frac{(\text{atomic}+\text{molecular})}{s^4}\sin\theta\Delta\phi\Delta\theta. \qquad \text{(B1)}$$

Hence, the fraction lost from the primary ray, for a *sample of infinitesimal thickness*,

$$\frac{dj(t)}{J_0}=2\pi\int_0^{\pi}\frac{j(\theta,t)d\theta}{J_0}=2\pi B^2N_0(t)\int_0^{\pi}\frac{(\text{atomic}+\text{molecular})}{s^4}\sin\theta d\theta\equiv\pi\sigma^2N_0(t). \qquad \text{(B2)}$$

Hence we have

$$\pi\sigma^2=2\pi B^2\left(\frac{\lambda}{2\pi}\right)^2\int_0^{4\pi/\lambda}\frac{(\text{atomic}+\text{molecular})_s}{s^3}ds. \qquad \text{(B3)}$$

The above integral may be evaluated through the use of Bewilogua's tables.* For CCl_4 this reduces to 1.71×10^{-17} cm², provided one makes appropriate assumptions as to the atom form factors at small angles. We also measured this quantity by making preliminary estimates of the magnitude of

$$\int_0^{R_0+\xi}N(R,t)dR \quad \text{and substituted in (A1).}$$

This leads to 2.1×10^{-17} cm² for the value of σ^2 for CCl_4.

We have thus demonstrated that the *fractional loss* due to higher order scattering is independent of the angle, and is directly measurable through the attenuation of the primary beam. For CCl_4 under conditions which are typical for most electron diffraction experiments, we estimated the average secondary and higher order scattering to be, after writing $N(R, t)=N_0(t)N_1(R)$,

$$\frac{1}{t_x}\int_0^{t_x}\frac{(I^*-I)}{I^*}dt=\frac{\pi\sigma^2}{t_x}\int_0^{t_x}N_0(t)dt\int^{R_0+\xi}N_1(R)dR=(4\text{ to }11)\times10^{-2} \qquad \text{(B4)}$$

* L. Bewilogua, Physik. Zeits. **32**, 740 (1931). See also W. Heisenberg, Physik. Zeits. **32**, 737 (1931).

where t_x is the exposure time. These are values which we believe are typical for most current electron diffraction experiments, and they certainly are greater than the error we wish to allow.

APPENDIX C. THE RATIO OF TOTAL MULTIPLE SCATTERING TO TOTAL SCATTERING

Divide interval 0 to $(R_0+\xi)$ into n slices of thickness $(\Delta R)_m$ so that each slice contains the same number of molecules, c/n, per unit area. Set $\pi\sigma^2=K$.

Then the amount of diffraction in the main beam occurring in slice m between R_m and R_{m+1} is

$$J_1(m)=J_0\left\{\exp\left[-K\int_{R_m+\Delta R_m}^{R_0+\xi}NdR\right]-\exp\left[-K\int_{R_m}^{R_0+\xi}NdR\right]\right\}$$
$$=J_0\left\{\exp\left[-K\frac{c(n-m)}{n}\right]-\exp\left[-K\frac{c(n-m+1)}{n}\right]\right\}.$$

The total amount of diffraction is therefore

$$J_1=\sum_{m=1}^{n}J_1(m)=J_0\{1-\exp[-Kc]\}.$$

From this we find at once that the total amount of intensity rescattered from the rays originating in slice m is given by*

$$J_2(m)=J_0\left\{\exp\left[-K\frac{c(n-m)}{n}\right]-\exp\left[-K\frac{c(n-m+1)}{n}\right]\right\}\cdot\left\{1-\exp\left[-K\int^{R_m}NdR\right]\right\}$$

and the total scattering from all rays originating in all the slices is

$$J_2=\sum_{m=1}^{n}J_2(m)=\sum_{m=1}^{n}J_0\left\{\exp\left[-K\frac{c(n-m)}{n}\right]-\exp\left[-K\frac{c(n-m+1)}{n}\right]\right\}\cdot\left\{1-\exp\left[-K\frac{m}{n}C\right]\right\}$$
$$=J_1+\sum_{m=1}^{n}J_0\left\{-\exp[-Kc]+\exp\left[-Kc\frac{n+1}{n}\right]\right\}$$
$$=J_1+nJ_0\{-\exp[-Kc]+\exp[-Kc(n+1/n)]\}$$
$$=J_1+nJ_0\{-1+\exp[-Kc/n]\}\exp[-Kc].$$

Now the value of

$$\lim_{n\to\infty}n\{e^{-Kc/n}-1\}=\lim_{n\to\infty}n\{-Kc/n\}=-Kc.$$

Hence

$$J_2=J_1+J_0\cdot(-Kc)\exp[-Kc]$$
$$=J_1+(J_0-J_1)(-Kc)$$
$$=J_1+(J_0-J_1)\ln[(J_0-J)/J_0]$$

or

$$J_2/J_1=1+[(J_0/J_1)-1]\cdot\ln[1-(J_1/J_0)].$$

* Ignores slight increase in path length because of angle and any change in gas density as a function of the distance *perpendicular* to the electron beam. These will be very small effects ordinarily since most of the intensity appears at small angles.

Studies with an Aerodynamically Instrumented Shock Tube*†

A. Hertzberg and A. Kantrowitz
Cornell University, Ithaca, New York

(Received April 6, 1950)

A shock tube has been set up at Cornell University to begin the study of two-dimensional wave propagation. Pressure waves initiated by the shock itself were used to trigger the spark source for the schlieren flow-observation system, to time the instant of exposure and to measure the strength (speed) of the shock waves. These techniques were employed to study the propagation of a wave in a converging channel. Fair agreement between the one-dimensional method of characteristics and these experiments was obtained.

INTRODUCTION

A SHOCK tube is a device in which simple normal shocks are produced by the sudden bursting of a diaphragm separating regions of higher and lower pressure gases. The compression waves produced in the lower pressure region rapidly steepen into a shock which propagates with a front normal to the axis of the channel. This device has proven convenient for many studies of shock wave phenomena.[1–5]

A shock tube has been constructed in the Graduate School of Aeronautical Engineering at Cornell and has been used for studies of the limitations of the one-dimensional characteristic theory.[6,7] Most previous shock tubes have used electronic instrumentation for measuring the strength of the shock wave generated and to trigger the spark at a desired time. The central idea in this work has been to simplify the instrumentation by using aerodynamic wave phenomena. Thus, weak waves traveling at nearly sonic speed were photographed simultaneously with the phenomena under study and

* This work has been largely supported by the ONR.
† Permanent equipment used in this research was furnished by the J. Carlton Ward Fund for Aeronautical Research of Cornell University.

[1] W. Payman and W. C. F. Shepherd, "Explosion waves and shock waves—The disturbances produced by bursting diaphragms with compressed air," Proc. Roy. Soc. A186, 293 (1946).

[2] F. B. Harrison and W. Bleakney, "Remeasurement of reflection angles in regular and Mach reflection shocks," Princeton University (1947).

[3] Huber, Fitton, and Delpino, "Experimental investigation of moving pressure disturbances and correlation with one-dimensional unsteady flow theory," N.A.C.A. T.N. 1903 (1949).

[4] C. Donaldson and R. D. Sullivan, "The effect of wall friction on the strength of shock waves in tubes and hydraulic jumps in channels," N.A.C.A. T.N. 1942 (September, 1949).

[5] G. R. Cowan and D. F. Hornig, "Technical report No. 1—The thickness of a shock front in air," Metcalf Research Laboratory, Brown University (April 6, 1949).

[6] A. Kantrowitz, "Heat engines based on wave processes" (Grad. Sch. of Aero. Eng.), Cornell University (1948). Presented at the annual meeting of the A.S.M.E. (November, 1948).

[7] R. Courant and K. O. Friedrichs, *Pure and Applied Mathematics Vol. 1, Supersonic Flow and Shock Waves* (Interscience Publishers, Inc., New York, 1948), p. 181.

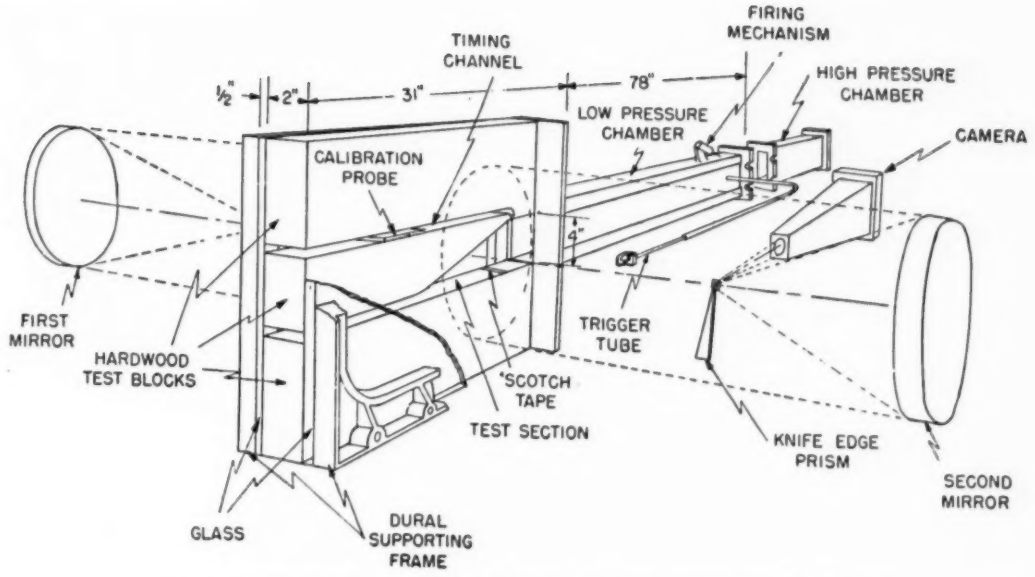


FIG. 1. Schematic drawing of shock tube and schlieren system used in this work.

were used to measure the shock strength and the time instant of the exposure.

APPARATUS

Standard Equipment

The shock tube used in these experiments (see Fig. 1) consisted of a rectangular steel tube 2×4 in. I.D. This tube was partitioned into high and low pressure chambers separated by a cellulose acetate diaphragm which could be pierced by a steel probe. By using various thicknesses of cellulose acetate up to 0.020 in., pressure differences between the chambers as large as 440 in. Hg could be used. In the calibration runs it was found that more consistent shock strengths were obtained if the diaphragm thickness was chosen to be close to its undisturbed bursting point when pierced. The end of the shock tube was equipped with a test section of interchangeable wooden blocks "sandwiched" between plate glass. A conventional schlieren apparatus was used to observe the flow. The spark source for the schlieren apparatus discharged 1.7 μf charged to about 9 kv. Exposures with a rotating mirror established the spark duration to be of the order of 1 μsec. A third trigger electrode was used which was energized via a thyratron.

Aerodynamic Instrumentation

The planned experimental program required the photography of shock waves by the schlieren system at various points during their propagation along the test section. Thus we needed to trigger the spark at controlled times and to measure the strength of the shock wave first for calibration purposes and later for the experimental work. The spark was triggered by a pressure wave which traveled down a small trigger tube (about $\frac{3}{16}$ in. I.D.—see Fig. 1). This trigger tube was connected to an orifice (about 2 in. from the diaphragm) and so received a pulse of high pressure air as the shock passed the orifice. Upon reaching the end of the trigger tube the pressure wave deflected a thin metal diaphragm a few thousandths of an inch to produce an electrical contact. This electrical contact was connected to the grid of the thyratron so as to trigger the spark. The trigger tube actually consisted of two metal tubes telescoping inside each other so that the spark could be triggered with appropriate lags to photograph the shock wave at various points along the test section with a reproducibility of about ± 10 μsec. Once one photograph of the shock wave was obtained for a given shock strength, pictures

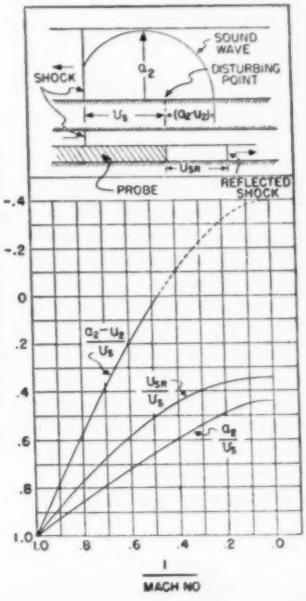


FIG. 2. Velocity ratios used for photographically measured shock strength by the application of weak waves (assumed infinitesimal in these calculations). The velocity ratio of a rearward moving weak wave $(a_2-u_2)/u_s$ was used where the shock strength was high and the velocity ratio of a normally reflected shock was used for measuring the strength of weak shocks, e.g., the timing wave.

FIG. 3. A schlieren photograph of reflections from a single thickness of Scotch Tape. The left edge of the vertical marker corresponds to the leading edge of the (invisible) Scotch Tape.

at other points in the test section could be obtained by extending or retracting the trigger tube approximately (within five percent) the distance between the new and old shock positions.

A more accurate method of recording the actual instant of exposure was through the use of the timing wave. In Fig. 1 it can be seen that a passage about $\frac{1}{2}$ in. high is included above the test section. This was connected to the main chamber by a narrow slit (0.015 in. wide and one-third the width of the channel). When the main shock passed this slit a very weak wave propagated along this upper channel at slightly greater than the speed of sound. The actual strength of this wave was measured with a calibration probe as will be discussed later. Its position at the instant of exposure was recorded along with the main shock wave on the schlieren photograph. Thus since the velocity of this wave is known from the calibration probe measurement (see below), its position is a measure of the time instant of exposure.

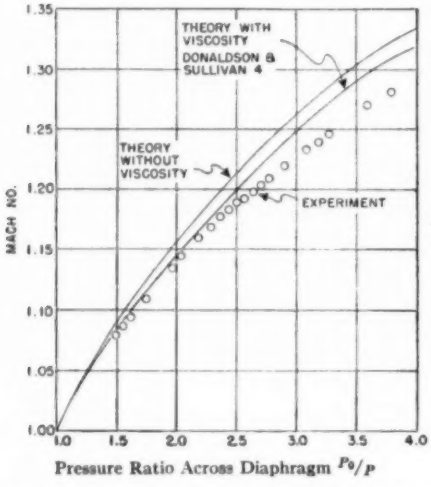


FIG. 4. Calibration of the shock tube. The experimental shock Mach numbers were obtained by the Scotch Tape technique (Fig. 3).

Shock Strength Measurement

Shock strengths were measured in all cases by the comparison of the speed of a reflected wave with the speed of a shock. The simplest technique is to place a piece of ordinary Scotch Tape on the floor of the shock tube in such a way that the passage of a normal shock will produce two weak cylindrical waves. The position of the upstream intersection of the rearward moving cylindrical wave with the wall and the position of a normal shock wave were measured relative to one edge of the Scotch Tape as shown in Fig. 2. The ratio of the distances of these waves from the disturbing point is identical with the ratio of the wave velocities. A photograph of the reflections from a piece of Scotch Tape is shown in Fig. 3. To measure the departure of these reflected waves from sound waves several different thicknesses of Scotch Tape were added to the original to produce stronger reflections. The apparent Mach numbers were plotted against tape thickness and the results extrapolated to zero thickness. No correction for a single thickness of Scotch Tape was required within the experimental scatter of the circles of the calibration on Fig. 4.

In order to calculate the strength of the shock from the ratio of the velocity of the reflected wave to the velocity of the normal shock wave, we need to know the strength of the shock wave in terms of u_2, a_2, and U_S where u_2 is the fluid velocity in the direction of the shock wave motion, a_2 is the velocity of sound behind the

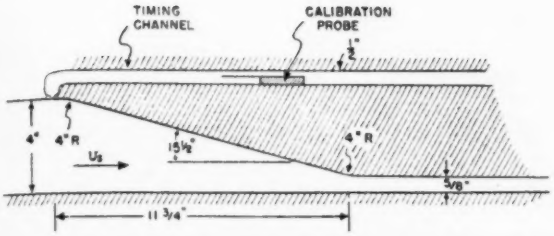


FIG. 5. Converging channel used in experimental work. The timing channel used to determine the exact time of the exposure is also illustrated.

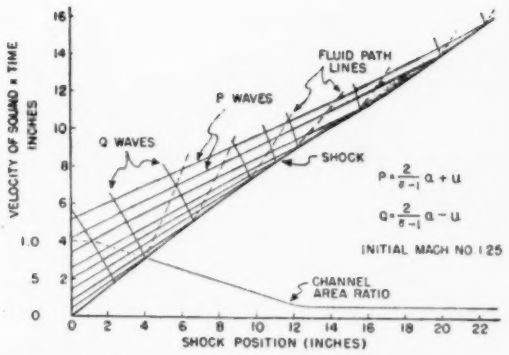


FIG. 6. Typical numerical integration along characteristics of the propagation of a shock wave (followed by no other disturbances) in a converging channel.

shock wave, and U_S is the velocity of the shock. It is desirable to transform to coordinates moving with the shock wave. Using this coordinate system the standard tabulations of the Rankine-Hugoniot relations can be used to give a relation between the shock Mach number and $(a_2-u_2)/U_S$. This relationship is plotted in Fig. 2. When the flow behind the shock becomes supersonic (the dashed curve on Fig. 2), this ratio would probably not be useful as simply measuring the Mach angle of the steady waves from the Scotch Tape. Similarly the wave moving across the channel can be used. In this case we measure the ratio a_2/U_S. This ratio is also plotted in Fig. 2. It can be seen that this ratio does not vary as rapidly with Mach number as the ratio between the upstream moving wave and the shock velocity. Therefore it does not lead to shock strength results as precise as those obtained from the upstream moving wave.

Normal reflection of the shock was used when the Scotch Tape reflection from a weak shock was no longer visible. A probe which is diagrammed in Fig. 2 was employed. In the lower part of this probe the wave was reflected normally whereas in the upper part it proceeded

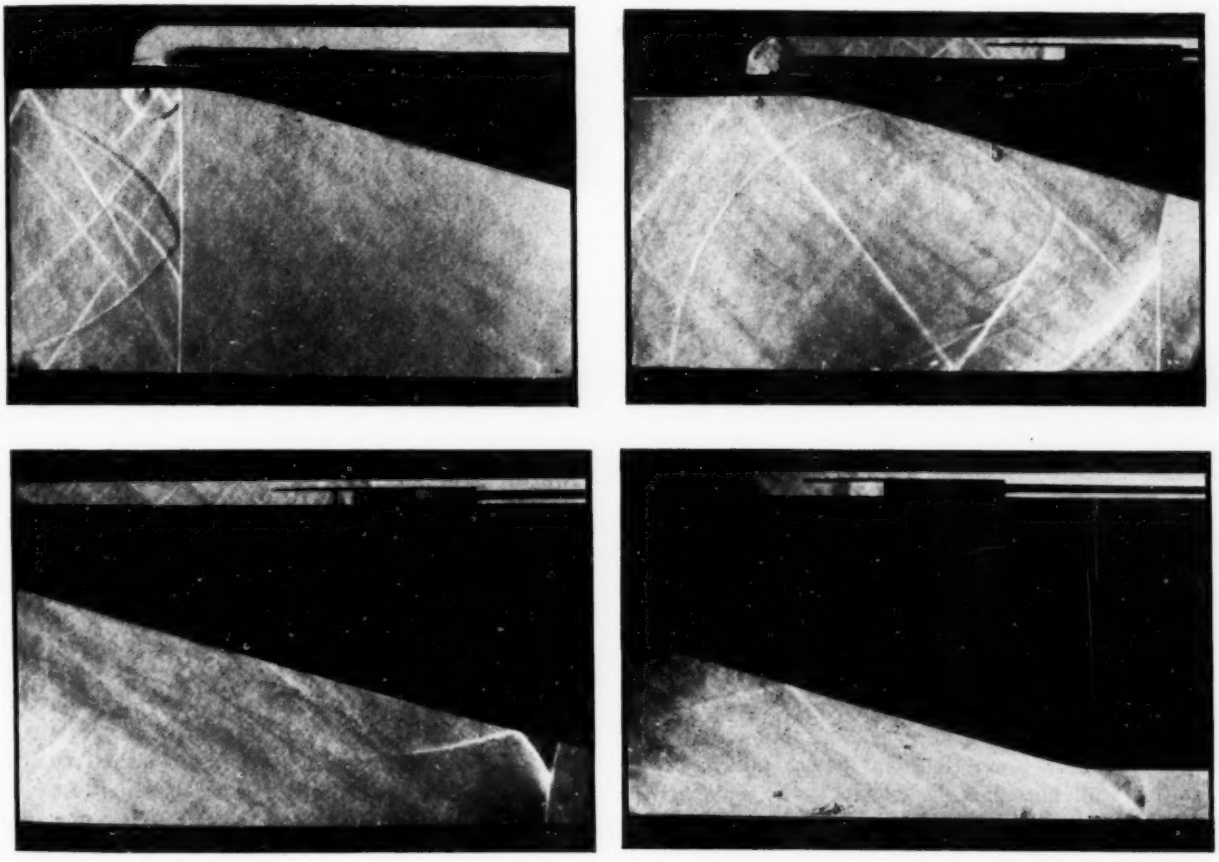


FIG. 7. Propagation of a shock (initial Mach No. 1.21) along the converging channel shown in Fig. 5. It will be observed [see Fig. 7(c)] that the shock rapidly assumes a cylindrical shape and then after a period of adaptation to the converging channel the one-dimensional theory of characteristics should apply.

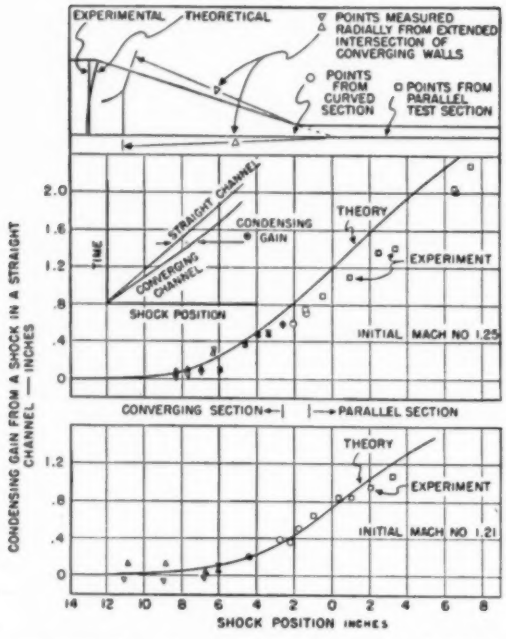


FIG. 8. Comparison of results obtained by numerical integration by the method of characteristics with experimental photographs such as Fig. 7.

with only slight disturbance. The ratio of the speed of the reflected wave to the speed of the initial wave can be used in a manner similar to the above for the determination of the shock strength. This ratio is also plotted in Fig. 2. The normal reflection technique was useful in finding the strength of the timing wave at various points along the timing channel, which had a Mach number of about 1.01.

Calibration of the Shock Tube

Before the shock tube can be conveniently used, a relation must be experimentally determined between the pressure ratio across the diaphragm and the Mach number of the resulting shock wave. This relationship can be determined from simple non-viscous theory by standard methods.[7] A viscous correction has recently been computed by Donaldson and Sullivan[4] which allows for the weakening of the shock wave by the boundary layer formed behind the shock wave. Both of these theoretical calibrations are plotted on Fig. 4. A series of photographs (such as Fig. 3) was taken at various pressure ratios using the Scotch Tape technique for measuring the shock strength. The measured shock Mach number is plotted against the pressure ratio on Fig. 4. It is seen that the Mach number obtained is always somewhat lower than the theory indicates even including wall boundary layer effects. This departure from the theoretical calculation including viscous effects might be expected in view of possible obstructions remaining at the diaphragm and miscellaneous losses due to partial reflections due to roughness in the shock tube and the test section.

APPLICATION TO THE PRELIMINARY STUDIES OF CONDENSING SHOCK WAVES

As was mentioned in the Introduction, our first experiments with the shock tube were directed to the study of the limitations of the one-dimensional method of characteristics for unsteady flows. It is assumed in this theory that conditions are uniform across the channel. On the other hand, if changes in the channel area occur, the effects will be felt first near the wall and only later in the center of the channel. In fact, it is clear from the photographs taken in the test channel that the effects of the wall led to a reflected wave similar to a triple shock intersection. This intersection point divides the shock into two portions, one of which has been affected by the change of wall contour and the other of which has not been affected by this change. We were also interested in determining whether a wave in a converging passage would ultimately assume a cylindrical form, i.e., to see whether the cylindrical form is stable for a condensing shock.

The channel chosen for these experiments is illustrated schematically in Fig. 5. The one-dimensional characteristic theory for this channel was computed by the methods of reference 6, and one such computation is shown in Fig. 6. No viscous effects are included in this calculation and no attempt to correct for departures from one-dimensional flow has been made.

The boundary conditions for the calculation were, first, that the flow is initially at rest with sonic velocity[8] a_o ($P=Q=[2/(\gamma-1)]a_0$) and, second, that no later waves propagated into the test section following the shock (i.e., at the beginning of the convergent section $P=[2/(\gamma-1)]a+u$ is a constant determined from the known initial shock strength).

A series of photographs was taken for two different shock Mach numbers (1.21 and 1.25) and typical photographs from the series for $M=1.21$ are shown in Figs. 7(a)–7(d). The earlier figures can be best described by saying that the effect of the wall deflection was propagated as a triple shock intersection back and forth across the channel. The later figures show that the shock does tend to approach a cylindrical form and thus indicate that the cylindrical form of the shock wave is stable in converging channels for the conditions of these experiments.

A comparison is given in Fig. 8 between the position of the shock wave as determined experimentally and the positions calculated by the method of characteristics. It will be noticed that the two ends of the shock approach the same radial position in about the time the triple shock intersection crosses the channel once.

The Magnetostriction of Permanent Magnet Alloys

E. A. NESBITT
Bell Telephone Laboratories, Murray Hill, New Jersey
(Received April 24, 1950)

In order to obtain a better understanding of the mechanism of coercive force in modern permanent magnets, magnetostriction measurements have been made on various alloys having coercive forces from 50 to 600 oersteds. The results can be summarized by discussing two types of alloys. First are the older carbon-hardening permanent magnets, and for these alloys high coercive force and high magnetostriction occur together. Second are the newer carbon-free permanent magnets and for these alloys high coercive force does not occur with high magnetostriction. In fact for the Mishima alloys having compositions near 29 percent nickel, 12.5 percent aluminum, and 58.5 percent iron, cooled at the rate of 3°C per second (coercive force 400 oersteds), the magnetostriction actually passes through zero. This is contrary to the classical strain theory of coercive force which states that the latter is proportional to the product of the magnetostriction and internal stress. To explain the mechanism of coercive force for these alloys it is necessary to resort to more recent theories.

INTRODUCTION

THE main purpose of this work has been to measure the magnetostriction of permanent magnet alloys having coercive forces from 50 to 600 oersteds, with the hope that the results would aid in understanding the mechanism of coercive force. At the present time the data on magnetostriction in this field are limited, although data of this type on materials of low coercive force are voluminous and have been of great value in interpreting the properties of materials of high permeability.

According to the older theories of Kersten,[1] Becker,[2] Kondorsky,[3] and others, a coercive force occurs when the material has (1) large internal strains that vary from place to place (strain gradients) and (2) large magnetostriction at saturation (change in length from demagnetized state to saturation). This classical strain theory of coercive force is expressed as follows:

$$H_c \doteqdot \lambda_s \sigma_i / I_s$$

where λ_s is the magnetostriction at saturation, σ_i is the amplitude of internal stress and I_s the saturation intensity of magnetization. Because of theoretical and experimental work, the validity of this theory has been questioned.

Néel[4] has developed a theory of coercive force in which he considers fluctuations in the magnetization of the elementary domains. Fluctuations in the *directions* of magnetization are caused by elastic disturbances of the lattice, and fluctuations in *intensity* of magnetization are caused by the presence of inclusions or cavities. Because of these fluctuations there will be magnetic poles throughout the material and associated with them will be energy, the distribution of which will largely determine the positions of the domain walls and their movement with a field.

Stoner and Wohlfarth[5] consider many materials having high coercive force (*magnets*) to be made basically of fine particles (this term including both actual powder grains and atomic segregates or islands in alloys) which act as single magnetic domains. The coercive force of these domains depends on one or more of three types of anisotropy, namely, strain, crystal and shape anisotropy. Strain anisotropy is a function of the strain and the magnetostriction. Crystal anisotropy depends on the crystal structure, and is measured by the anisotropy constant. Shape anisotropy depends upon the shape of the particle in such a way that it is zero for

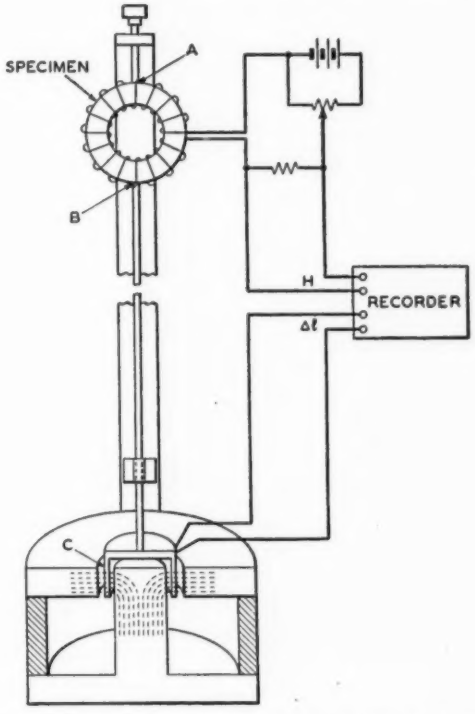


FIG. 1. Schematic diagram of magnetostriction instrument. Fluxmeter may be used in place of recorder.

[1] M. Kersten, Zeits. Tech. Physik **12**, 667–9 (1931).
[2] R. Becker, Physik. Zeits **33**, 905–13 (1932).
[3] E. Kondorsky, Physik. Zeits Sowjetunion (U.S.S.R.) **8**, 153 (1937).
[4] L. Néel, Comptes Rendus **224**, 1488 (1947).
[5] Stoner and Wohlfarth, Phil. Trans. Roy. Soc. (London) **240**, 599–644 (1948).

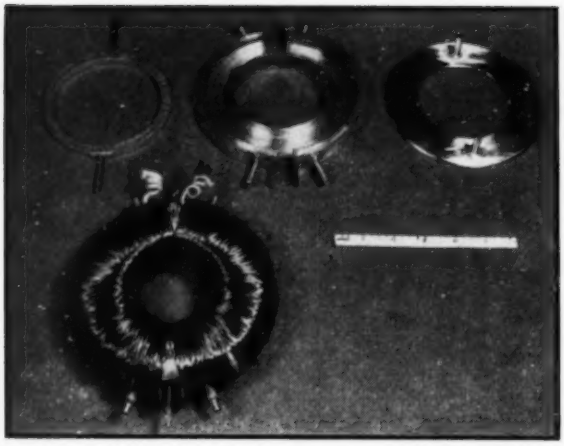


FIG. 2. Photograph of specimen, box, and winding details.

a material having spherical particles and high for a material composed of long, thin needles. Magnets composed of fine particles of this kind, and having high coercive force, have been produced by Dean and Davis[6] (powdered iron, cobalt, etc.), Guillaud[7] (powdered MnBi, MnAs, etc.), Weil,[8] and others. The theory has been discussed by Néel, and Stoner and Wohlfarth and by Kittel.[9] In the case of iron, the particles must be of the order of 2×10^{-6} cm or smaller in order to act as single magnetic domains.

The measurements here reported make it necessary in the case of some important alloys to resort to the new theories of Néel, and Stoner and Wohlfarth and to reject the older strain theory. For example in some of the Mishima alloys zero magnetostriction was obtained with an accompanying high coercive force and this is incompatible with the strain theory. In other materials such as high cobalt steel the strain theory may still apply.

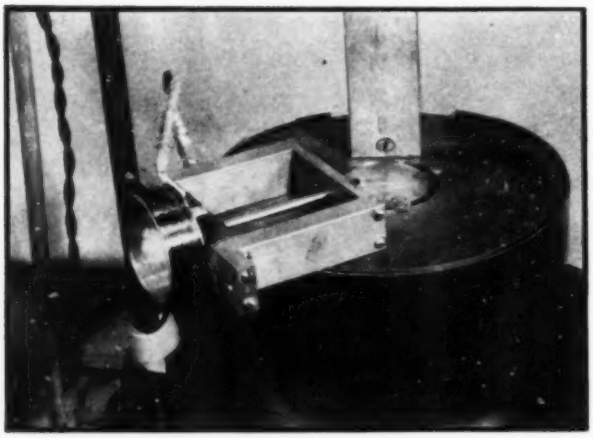


FIG. 3. Photograph of cylindrical coil, bearings, and circular gap in magnet.

[6] Dean and Davis, U. S. Patent 2,239,144.
[7] Guillaud, Thesis, Strasbourg (1943).
[8] L. Weil, Comptes Rendus **225**, 229 (1947).
[9] C. Kittel, Phys. Rev. **70**, 965 (1946).

APPARATUS

The first phase of this work was the development of a new method for measuring magnetostriction which was originally devised by H. J. Williams. The method is a magnetic one and has the desirable feature of recording automatically the change in length of a specimen as a function of the magnetizing force. Figure 1 shows a schematic diagram of this instrument.

The toroidal specimen is rigidly clamped to the brass frame at the point A as shown in Fig. 1. At a point

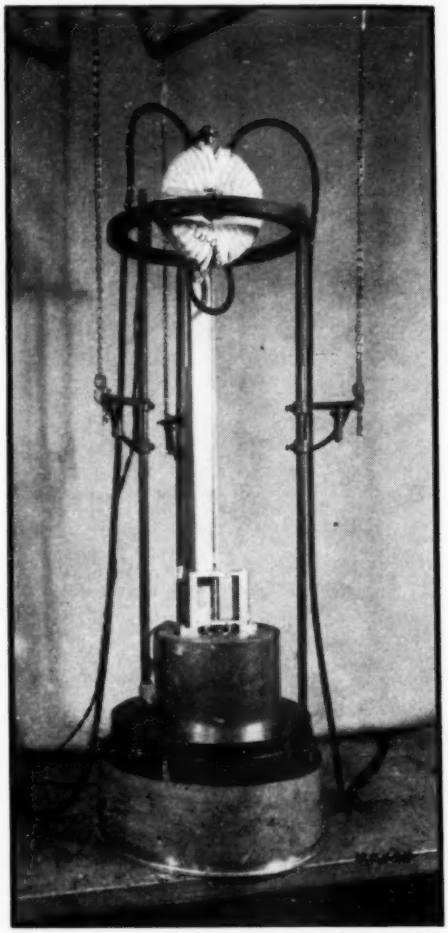


FIG. 4. Photograph of complete assembly of magnetostriction instrument.

diametrically opposite, a brass rod is soldered to the specimen and this rod supports a cylindrical coil C which is suspended in a uniform radial magnetic field supplied by permanent magnets. Any vertical displacement of the cylindrical coil causes a change in the flux linkages of the coil and this is measured by the recorder. The recorder is a Cioffi recording fluxmeter which consists of two elements, one drives a paper drum proportional to the magnetizing force, the other drives the pen across the paper in proportion to the change in flux linkages. In practice the instrument is actuated by

sending current through the magnetizing winding of the specimen. This causes two things to happen. First, sending current through the magnetizing winding also sends current through the recorder and this causes the paper drum to turn by an amount proportional to the magnetizing force. Second, sending current through the magnetizing winding magnetizes the specimen and changes its diameter thus causing a vertical displacement of the cylindrical coil and as a result the pen then moves an amount proportional to the change in flux linkages. Both of these actions occur simultaneously, thus a curve is traced of the change in length of a specimen *versus* magnetizing force.

Figure 2 shows the specimen, the box in which the specimen is suspended but not allowed to touch, the cover for the box and the completely wound specimen. The specimen shown is approximately 4 in. I.D., 4.6 in. O.D., and is either a cast or a machined ring. Specimens 2 in. I.D., 2.5 in. O.D. were also used. The brass extensions are soldered to the specimen, the heavy one fastens to the brass frame, the light one supports the

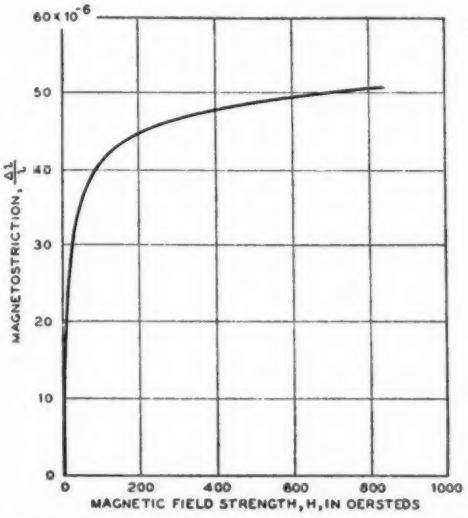


FIG. 5. Longitudinal magnetostriction of Vanadium-Permendur.

cylindrical coil. The box and its cover are watercooled; this is important because it is necessary to keep the temperature of the specimen constant during the measurement. The specimen shown in Fig. 2 is wound with 1850 turns of No. 18 Formex wire. The smaller specimens were wound with approximately 800 turns of No. 19 enameled silk copper wire. It is important to make the windings uniform and continuous in order to avoid magnetic pick-up between the specimen and the cylindrical coil. In some cases it is desirable to have a cylindrical magnetic shield surrounding the toroidal magnetizing winding.

Figure 3 shows the cylindrical coil and its associated bearings for guiding the motion of the coil. The coil is wound with 1600 turns of No. 38 enameled copper wire. The bearings are fitted to have a minimum of frictional

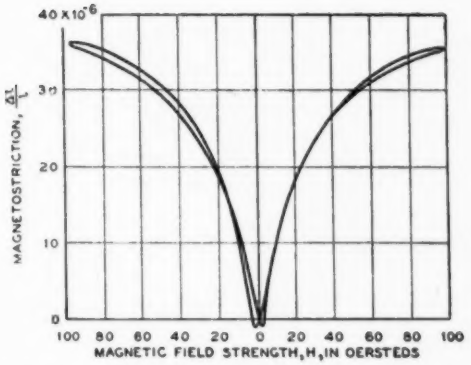


FIG. 6. Magnetostrictive loop of Vanadium-Permendur.

resistance and yet prevent side play of the coil *C*. The coil is placed in the circular gap.

Figure 4 shows the complete assembly of the instrument. The instrument is mounted on a base which is supported by a Julius suspension, part of which can be seen in the photograph. The base is oil damped and the entire instrument enclosed in an air shield (not shown in photograph) since it is important to reduce all mechanical and wind vibrations to a minimum. The rubber tubes shown in the photograph carry water to cool the box containing the specimen. The distance between the specimen and the coil should be at least two feet.

In order to test the accuracy of this new instrument several of the soft magnetic materials were measured since these are freer from cracks, easier to magnetize and chemically more homogeneous. The latter factor is important in measuring magnetostriction because any inhomogeneity in the specimen causes a pole to occur and this results in an undesirable extraneous force between the specimen and the magnetizing coil. Figure 5

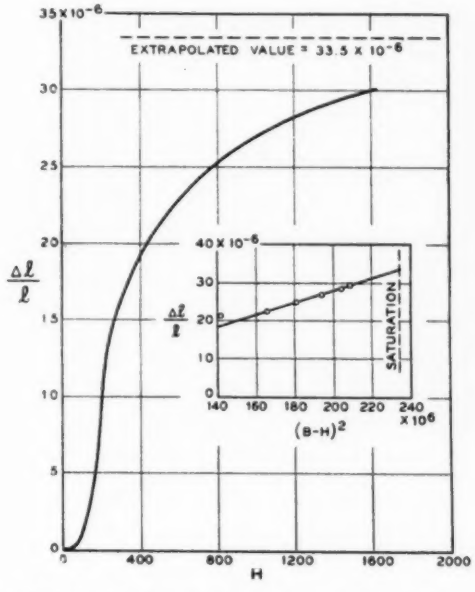


FIG. 7. Magnetostriction of Alnico V after cooling from 1300°C at 2°C per second.

TABLE I. Magnetostriction of Alnico V.

Treatment	Size	$\frac{\Delta l}{l}10^6$ $H=1600$	$\frac{\Delta l}{l}10^6$ maximum	Magnetic properties B_r	H_c	B_m	H for B_m
1300°C—cooled approximately 2°C per second	4 in. I.D. 4.5 in. O.D.	30	33.5	11650	270	16300	1510
1300°C—cooled approximately 2°C per second		27.5	35	10900	140	16420	1920
1300°C—cooled approximately 2°C per second plus 585°C for 8 hr.	2 in. I.D. 2.5 in. O.D.	21.8	22.3	9200	450	15150	2180
1300°C—cooled approximately 2°C per second to 820°C then oil quenched		26	26	9060	13	15300	1120
Sand casting	4 in. I.D. 4.5 in. O.D.	19	22.3	9620	201	15100	1940
1000°C—1 hr.–furnace cooled	2 in. I.D. 2.5 in. O.D.	17.2	25	8550	298	14900	1930
1300°C—cooled approximately 2°C per second in a transverse magnetic field	2 in. I.D. 2.5 in. O.D.	36	43	5140	154	15220	1900
1300°C—cooled approximately 2°C per second in a longitudinal magnetic field	2 in. I.D. 2.5 in. O.D.	3.5		12900	252	16200	1460
Cooled in longitudinal field, quenched at 820°C	2 in. I.D. 2.5 in. O.D.	2.5	2.5	13250	12.8	14500	147
Cooled in transverse field, quenched at 820°C	2 in. I.D. 2.5 in. O.D.	38	38	610	17.2	15400	895

TABLE II. Magnetostriction of alloys in the vicinity of the Alnico V composition.

Composition					Treatment	Ring size	$\frac{\Delta l}{l}10^6$ $H=1600$	$\frac{\Delta l}{l}10^6$ maximum	B_r	H_c	B_m	H for B_m
Ni	Al	Co	Cu	Fe								
14	2	24	3	57	1000°C—1 hr.–F.C.	2.0 in. I.D. 2.5 in. O.D.	30.4	38	5850	18	21600	1160
14	6	24	3	53	1300°C—cooled 2°C per second to 820°C then oil quenched		19.2	19.2	4010	22.4	17100	1120
14	6	24	3	53	Sand cast	4.0 in. I.D. 4.5 in. O.D.	25.5	25.6	4540	56.6	17000	1700
14	6	24	3	53	1300°C—cooled approx. 2°C per second	4.0 in. I.D. 4.5 in. O.D.	25.3	25.3	7120	86	17000	1640
14	6	24	3	53	1000°C—1 hr.–F.C.	2.0 in. I.D. 2.5 in. O.D.	14.2	14.4	6990	103	17100	1570
14	10	24	3	49	Sand cast	1.9 in. I.D. 2.5 in. O.D.		19.0	9330	4.0	15500	1590
14	10	24	3	49	1000°C—1 hr.–F.C.	2.0 in. I.D. 2.5 in. O.D.	14.2		7570	249	16300	1900

TABLE III. Magnetostriction of alloys of varying coercivity.

Composition	Treatment	Ring size I.D.	$\frac{\Delta l}{l}10^6$ $H=1600$	$\frac{\Delta l}{l}10^6$ maximum	B_r	H_c	B_m	H for B_m
Tool steel (0.9 percent carbon)	800°C—5 min.—brine quench +100°C—48 hr.	2 in.	−2.8		10850	45	19100	1250
Low cobalt steel (0.9C—4.75Cr—1.25W 8.5Co—bal. Fe)	900°C—5 min.—oil quench +100°C—24 hr.	2 in.	+14.5	+14.5	9000	93	16720	1340
Honda steel (0.8C—4W—2Cr—36Co—bal. Fe)	940°C—7 min.—oil quench +100°C—24 hr.	2 in.	+33.5	+40.5	10180	183	19900	1860
Remalloy (71Fe—12 percent Co—17Mo)	1300°C—20 min.—oil quench	2 in.		24	4270	2.6	17600	354
Remalloy (71Fe—12 percent Co—17Mo)	1300°C—20 min.—oil quench +685°C—1.5 hr.	4 in.	+11.8	+17	9580	305	16400	1860
Vicalloy (37.5Fe—52Co—10.5V)	1000°C—30 min.—brine quench	2 in.	+34.2	+47.5	4900	77	16300	2100
Vicalloy (37.5Fe—52Co—10.5V)	1000°C—30 min.—brine quench +600°C—1 hr.	2 in.	+26.0	+34	7500	320	14100	1660
Alnico V	1300°C—20 min.—cooled 2°C per second	4 in.	+30	+33.5	11650	270	16300	1510
Alnico V	1300°C—20 min.—cooled 2°C per second	4 in.	+27.5	+35	10900	140	16420	1920
Alnico V	1300°C—20 min.—cooled 2°C per second +585°C—8 hr.	2 in.	+21.8	+22.3	9200	450	15150	2180
Cunico (21Ni—50Cu—29Co)	Commercial heat treatment	2 in.	+3.7	+7.5	4140	445	8820	2400
Vicalloy tape (38Fe—52Co—10V)	800°C—1 min.—H_2 quench	2 in.	+10.5	+11	1700	195	7150	1900
23Ni—12Al—65Fe	1300°C—30 min.—cooled 5°C per second	2 in.		−1.0	6800	356	12800	1960
29Ni—13Al—58Fe	1300°C—30 min.—cooled 5°C per second	2 in.	−3.0	−6.2	5000	482	10880	1965
29Ni—13Al—58Fe	1300°C—30 min.—cooled 5°C per second +650°—2 hr.	2 in.	−5.5	−10.5	5080	461	10850	1730
29Ni—13Al—58Fe	As cast in sand	2 in.	−5.9	−15.0	4850	560	11320	2040
29Ni—13Al—58Fe	1000°C—1 hr.–F.C.	2 in.	−7.0	−17.5	2700	134	12200	2400
29Ni—10Al—63Fe	1350°C—cooled 4°C per second	2 in.	+7.0	+13	5050	340	10530	1720
29Ni—12Al—61Fe	1350°C—cooled 4°C per second	2 in.	+3.6	+5.5	4640	420	10700	1945
29Ni—12.5Al—60.5Fe	1350°C—cooled 4°C per second	2 in.	−2.0	−4.5	4400	400	11180	2410

shows an actual magnetostrictive trace obtained with this instrument in conjunction with the Cioffi recording fluxmeter, using a specimen of Vanadium-Permendur, an alloy of 49 percent iron, 49 percent cobalt, and 2 percent vanadium. At a field strength of 830 oersteds, a magnetostriction of 51×10^{-6} was obtained. The specimen was approximately 4 in. I.D., 4.5 in. O.D., and had been annealed at 1000°C for 1 hour and slowly cooled. This agrees closely with values obtained by other investigators, for example, on this alloy not containing the vanadium, Honda[10] obtained a magnetostriction of approximately 50×10^{-6} with $H=500$ oersteds. Sometimes it is desirable to take magnetostrictive loops over limited field strengths and Fig. 6 shows that this

[10] K. Honda, *Magnetic Properties of Matter* (Tokyo, 1928), p. 111.

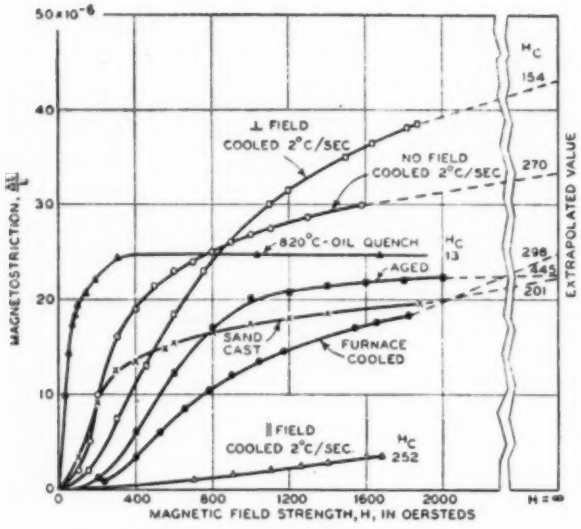


FIG. 8. Magnetostriction of Alnico V for seven different treatments.

instrument is capable of taking symmetrical loops of this type. The specimen is the same as the one used for Fig. 5. The specimen was first magnetized at its remnant value and the magnetizing force then applied in the opposite direction. The magnetostriction dipped slightly then rose to a maximum value of 35.6×10^{-6}, and then as the H was decreased, the magnetostriction decreased to zero. The H was reversed, the magnetostriction dipped again and then rose to a maximum of 36.2×10^{-6} and then as the H was decreased again, the magnetostriction decreased to zero, its starting point.

In many of the experimental curves, the saturation magnetostriction was not reached since magnetizing forces of only about 2000 oersteds were used. It is difficult to saturate completely permanent magnets in their hardened state, and this is especially true when toroidal specimens are used. The use of toroidal specimens has several advantages, however, because by this method one avoids as much as possible the translational forces caused by non-uniform fields and also avoids errors resulting from the forces between free poles. Figure 7 shows trace of the magnetostriction as a function of the magnetizing force for Alnico V and an extrapolated curve (data obtained from trace) of magnetostriction as a function of $(B-H)^2$ for obtaining the saturation magnetostriction. Sometimes it was also necessary to plot $(B-H)$ *vs.* $1/H$ to obtain the saturation magnetization by extrapolation.

The measurement of zero magnetostriction on a polycrystalline specimen does not necessarily mean that the magnetostrictions in the various crystallographic directions are zero. It is possible that a negative magnetostriction in the [111] direction could be balanced somewhat by two positive magnetostrictions in the [100] and [110] directions as is the case in iron. However, an exact balance of these crystallographic magnetostrictions for all values of flux density is unlikely. When the polycrystalline magnetostriction is found to be low at all inductions, it is believed that the magnetostriction is low in all crystallographic directions.

RESULTS

The data obtained may be divided into four categories.

(1) The magnetostriction of Alnico V and its variation with field and heat treatment.

(2) The magnetostriction of alloys having compositions near that of Alnico V.

(3) The magnetostriction of permanent magnet alloys of varying coercivity. The purpose of these experiments is to note any general relationship between magnetostriction and coercive force.

(4) The magnetostriction of several alloys of special interest which cannot be classified as permanent magnets. A specimen of Ferroxcube received from the American Philips Company, through the kindness of Dr. F. Brockman, and specimens of oriented iron-cobalt were measured. An iron-cobalt specimen yielded a magnetostriction of 130×10^{-6} one of the highest values ever reported.

The data for the first three parts are summarized in Tables I–III and they will now be discussed in detail.

Alnico V is probably the most important permanent magnet material. This alloy and associated alloys which are modifications of Alnico V are the only commercial permanent magnets which respond to heat treatment in

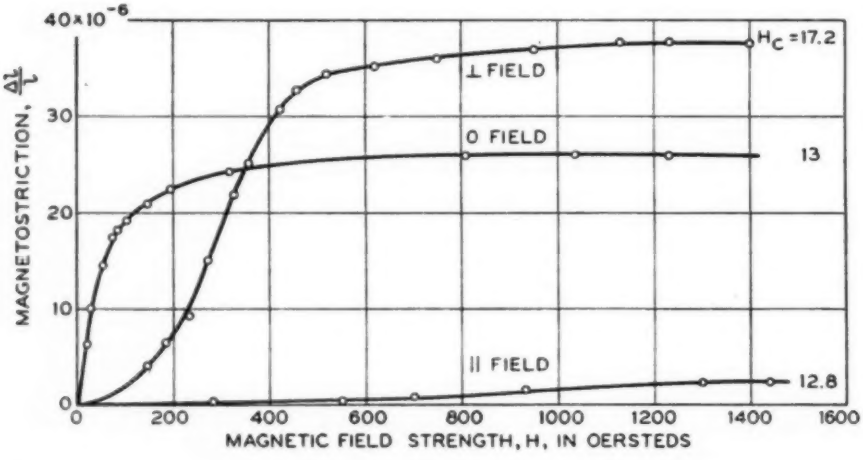


FIG. 9. Magnetostriction of Alnico V quenched from 820°C.

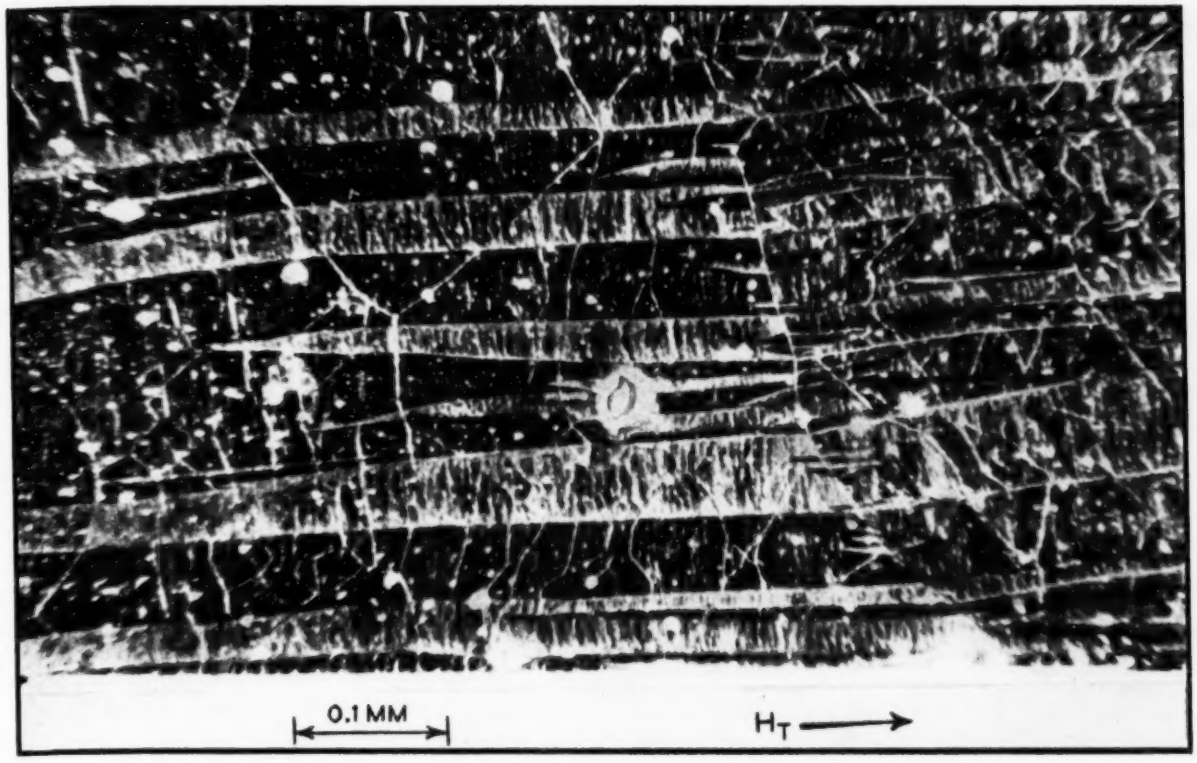


FIG. 10. Powder pattern photograph of Alnico V heat treated in a longitudinal magnetic field, the direction of which is shown by the arrow.

a magnetic field. Suggestions[11] have been made for explaining its response to heat treatment in a field and the nature of its hardening mechanism, and these suggestions have been aided by measurements of magnetostriction.

Figure 8 shows the magnetostriction of Alnico V as a function of magnetizing force for seven different treatments. The saturation magnetostriction values of the quenched, sandcast, and furnace cooled states are respectively 26×10^{-6}, 22.3×10^{-6}, and 25×10^{-6}. In the quenched specimen the precipitation process has not yet begun, in the sandcast specimen the process has not yet advanced to the point for optimum magnet properties while in the furnace-cooled specimen the process has advanced beyond the optimum point. The optimum cooling rate for Alnico V is approximately 2°C per second from 1300°C and the curve for zero field shows the magnetostrictive properties obtained on a specimen cooled at this rate. At an H of 1600 oersteds, the magnetostriction was 30×10^{-6} and the extrapolated value for saturation was 33.5×10^{-6}. The magnetic properties of this specimen are residual induction (B_r) of 11650 gauss, coercive force (H_c) of 270 oersteds, and the B for an H of 1510 oersteds is 16300 gauss. Another specimen of Alnico V (not shown in figure) heat treated in a similar manner had a magnetostriction of 27.5×10^{-6} at an H of 1600 oersteds and an extrapolated value of 35×10^{-6}. These two specimens cooled at the optimum rate yielded the highest magnetostriction for specimens of Alnico V not heat treated in a magnetic field. When Alnico V is completely hardened the saturation magnetostriction is lowered, a value of 21.8×10^{-6} was obtained at an H of 1600 oersteds the extrapolated value was 22.3×10^{-6} as shown in the curve for the aged specimen. The magnetic properties of this specimen were $B_r=9200$ gauss, $H_c=450$ oersteds and a saturation magnetization of 14150 gauss. The results given so far are for Alnico V not heat treated in a magnetic field and a random domain distribution was believed to exist. Specimens of Alnico V heat treated in longitudinal and transverse magnetic fields were also measured. The magnetostriction of the specimen heat treated in a longitudinal field was low as would be expected from the domain theory. A value of only 3.5×10^{-6} at an H of 1680 oersteds was obtained as shown in the curve for the parallel magnetic field. The magnetic properties of this specimen were $B_r=12900$ gauss, $H_c=252$ oersteds and $B=16200$ gauss for an H of 1460 oersteds. The magnetostriction of the specimen heat treated in a transverse field was high as would be expected from domain theory. A value of 36×10^{-6} was obtained at an H of 1600 oersteds and the extrapolated value was 43×10^{-6}. This is higher than that obtained on the specimens cooled at the same rate but with no applied field during the heat treatment. The magnetic properties of the

[11] Kittel, Nesbitt, Shockley, Phys. Rev. **77**, 839–840 (1950).

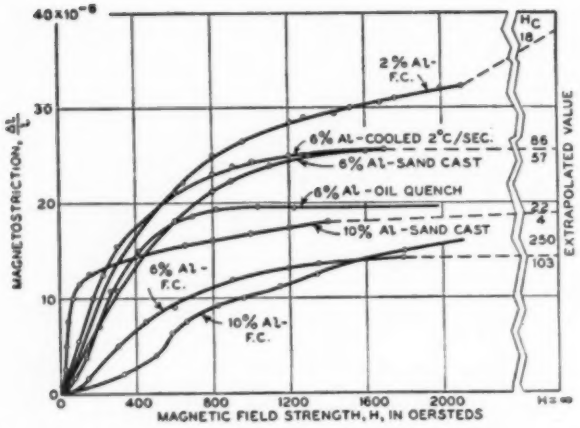


FIG. 11. Magnetostriction of several alloys having compositions near Alnico V. F. C. indicates furnace cooled.

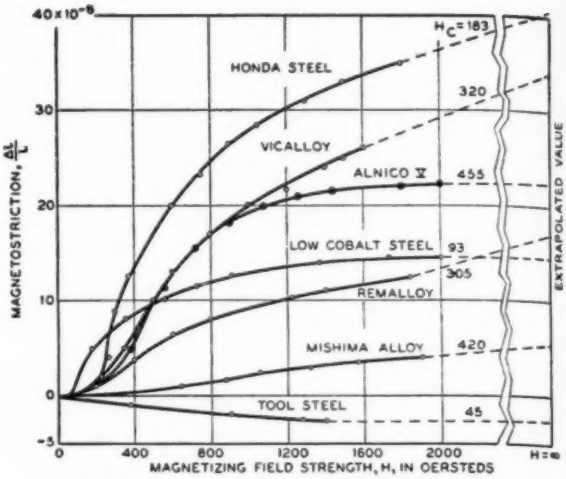


FIG. 12. Magnetostriction of permanent magnet alloys of different coercivity.

specimen were $B_r = 5140$ gauss, $H_c = 154$ oersteds, and $B = 15220$ gauss for an H of 1900 oersteds.

Figure 9 compares the magnetostriction of three specimens of Alnico V which were quenched from 820°C. The first specimen was cooled in a transverse field and its magnetostriction was 38×10^{-6}. This value is almost one and one-half times higher than the magnetostriction of the second specimen which was cooled in zero field. This is in exact accordance with domain theory which predicts that a specimen having the domains oriented transversely will have one and one-half times the magnetostriction of a specimen having random domain orientation. The third specimen was heat treated in a longitudinal field and its magnetostriction was 2.5×10^{-6}. This low value is obtained because the longitudinal field orients the domains either parallel or antiparallel to the direction of the applied field and so very little change in length is observed upon magnetization. A powder pattern photograph of the domains in this ring is shown in Fig. 10, taken with H. J. Williams. These large domains or domain aggregates extend around the circumference of the ring in the same direction that the field was applied during heat treatment. As a result of this phase of the investigation, we may say that Alnico V normally has a high magnetostriction and that it varies greatly with heat treatment from 2.5×10^{-6} to 43×10^{-6}. The behavior of the specimens heat treated in a magnetic field conforms with modern domain theory.

The second phase of this work consisted of measuring the magnetostriction of several alloys having aluminum contents in the vicinity of standard Alnico V (8 percent aluminum) to note how the magnetostriction varies with composition. The sum of the aluminum and iron percentages was kept constant. Figure 11 shows seven curves of magnetostriction as a function of magnetizing force for alloys of two, six, and ten percent aluminum. The magnetostriction of the six percent aluminum alloy for sandcast, cooled 2°C per second, quenched, and furnace cooled states, are respectively 25.5×10^{-6}, 25.3×10^{-6}, 19.2×10^{-6}, and 14.2×10^{-6} for an H of 1600 oersteds. With this alloy as with the standard 8 percent aluminum, the magnetostriction depends on heat treatment. Jellinghaus[12] mentioned that alloys containing 5 to 7 percent aluminum have a high Curie point but for some unknown reason fail to respond to heat treatment in a magnetic field. We now know that the 6 percent aluminum alloy also has a substantial value of magnetostriction although its value is only 75 percent that of the Alnico V. Alloys containing only 5 to 7 percent aluminum develop only a fraction of the coercive force of Alnico V. Magnetostriction values of the 10 percent aluminum alloy for the sand cast and furnace cooled states are respectively 17.5×10^{-6} and 12.9×10^{-6} for an H of 1400 oersteds. The furnace cooled specimen has an irregular curve indicating a heterogeneous structure. The magnetization curve has a similar irregularity showing that the magnetostriction is following the magnetization. The magnetostriction of the furnace cooled alloy with 2 percent aluminum was 30.4×10^{-6} at an H of 1600 oersteds and 38×10^{-6} at saturation. This alloy containing only 2 percent aluminum is not in the immediate vicinity of Alnico V and its magnetostriction is due primarily to the magnetic constituents of iron, nickel and cobalt and not to the aluminum.

The third phase of this work consisted of measuring the magnetostriction of a number of permanent magnets having different values of coercive force, as shown in Fig. 12. For tool steel at an H of 1600 oersteds the magnetostriction has a value of approximately -2.8×10^{-6}. This low negative value is also accompanied by a low value of coercive force (45 oersteds). The addition of cobalt to steel (along with several carbide producing elements such as tungsten and chromium) produces a series of permanent magnets of increasing coercivity. The magnetostriction of a low cobalt steel (8.5 percent cobalt) is shown in Fig. 12 and has a value of 14.3×10^{-6}

[12] W. Jellinghaus, Arch. Eisenhüttenw. 16, 247–251 (1943).

at $H=1600$ and a coercive force of 93 oersteds. The best of these cobalt steels is Honda steel and its magnetostrictive curve shown in Fig. 12 yields a value of 33.6×10^{-6} at an H of 1600 oersteds, accompanied by a coercive force of 183 oersteds. Thus we see that in the cobalt steel system for the alloys measured as the magnetostriction increases the coercive force does likewise. This result is consistent with the classical strain theory of coercive force (coercive force is proportional to the product of magnetostriction and stress) if we assume that the stress is substantially constant in these alloys. This seems like a reasonable assumption since the alloys are closely related to each other and their indentation hardness is approximately the same.

Remalloy,[13] a hot workable alloy of iron, cobalt, and molybdenum, and Vicalloy,[14] a hot or cold workable alloy of iron, cobalt, and vanadium, were measured. The magnetostriction of the Remalloy specimen as 11.6×10^{-6} at 1600 oersteds, and the extrapolated value was 17×10^{-6} (Fig. 12). The magnetostriction of the Vicalloy specimen was 26×10^{-6} at 1600 oersteds and the extrapolated value was 34×10^{-6}. The coercive force of the Remalloy was 305 oersteds, the coercive force of the Vicalloy was 320 oersteds while the respective magnetostrictions were 17 and 34×10^{-6}. In this instance, therefore, we have two different alloys with approximately the same coercive force but one alloy has twice the magnetostriction of the other. In addition the magnetostrictions of the Remalloy and Vicalloy specimens were measured in an unhardened state and the values obtained were respectively 24 and 47.5×10^{-6} (Table III). These values are considerably higher than those obtained on the hardened specimens, it being generally true that age-hardening lowers the magnetostriction of an alloy.

The magnetostriction curve for the hardened Alnico V (not heat treated in a field) is replotted in Fig. 12. The magnetostriction was 21.8×10^{-6} at 1600 oersteds and the extrapolated value was 22.3×10^{-6}, and the coercive force was 450 oersteds in the same specimen.

The magnetostriction of the Mishima alloy (29 percent Ni, 12 percent Al, 59 percent Fe), shown in Fig. 12, was so low for its accompanying coercive force of 420 oersteds that it was decided to investigate this system in detail. A series of Mishima alloys were made with nickel held constant at 29 percent and aluminum varying from 8 to 13 percent, and the results obtained are shown in Fig. 13. The magnetostriction of these alloys gradually diminishes with increasing aluminum content and actually passes through zero in the vicinity of 12 to 12.5 percent aluminum. According to the strain theory of coercive force, the coercive force should also pass through zero in this vicinity but the alloys in this range have coercive forces from 400 to 440 oersteds. Evidently the strain theory does not apply to these alloys and in all probability it does not apply to Alnico V since this alloy is directly related to the iron-nickel-aluminum system. The alloys were cooled from 1350°–1300°C at the rate of 2°–5°C per second and their magnetostriction depends to a certain extent upon this heat treatment. This is shown in Table III for the alloy containing 13 percent aluminum. Saturation magnetostriction values for the annealed, cast, aged, and cooled at 5°C per second states are respectively -17.5×10^{-6}, -15.0×10^{-6}, -10.5×10^{-6}, and -6.2×10^{-6}.

A number of the alloys having coercive forces from 50 to 600 oersteds are shown graphically in Fig. 14, in which coercive force is plotted against magnetostriction. In addition to the results shown in Figs. 12 and 13, magnetostriction is plotted for the alloy Cunico (50 percent Cu, 21 percent Ni, 29 percent Co), Vicalloy tape of low magnetic saturation, and two Mishima alloys. Information on the indentation hardness and magnetic saturation are also included. The indentation hardness probably bears some relation to the amplitude of the internal stress. Also in a given material, if all other factors remain the same one would expect the magnetostriction to depend directly on the magnetic saturation. The Cunico specimen has a low value of magnetostriction (7.5×10^{-6}), a low indentation hardness (R.C. 20) and a high coercive force (445 oersteds) and it therefore contributes strongly to the idea that the classical strain

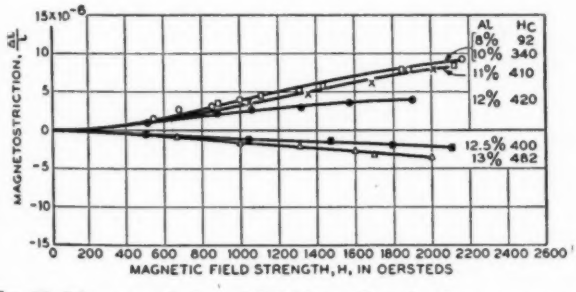


FIG. 13. Magnetostriction of Mishima alloys having a constant nickel content of 29 percent.

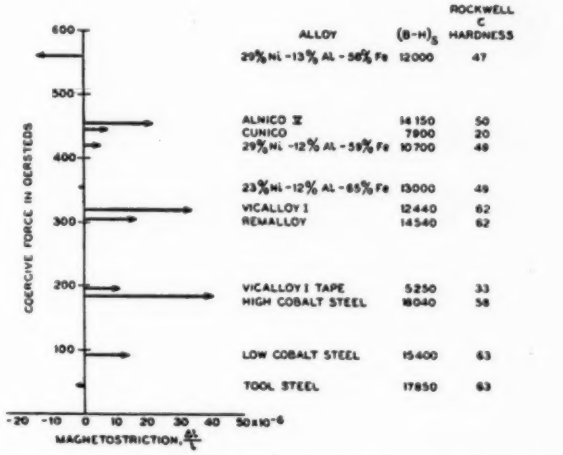


FIG. 14. Magnetostriction of alloys having coercive forces in the range of 50 to 600 oersteds.

[13] V. E. Legg, Bell Sys. Tech. J. **18**, 248 (1939).
[14] E. A. Nesbitt, Trans. Am. Inst. Min. Met. Engrs. 166 (1946).

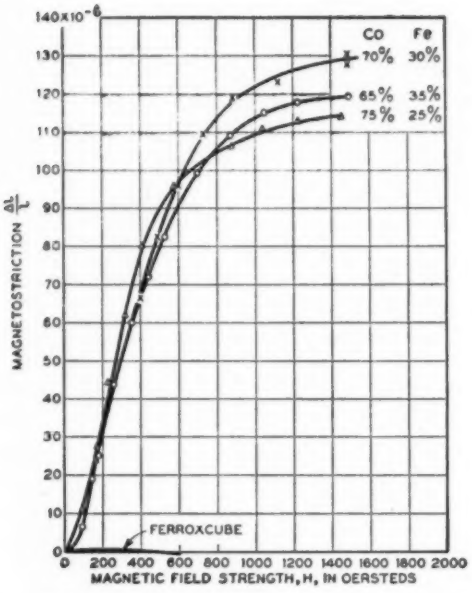


FIG. 15. Magnetostriction of alloys of special interest.

theory does not apply to this alloy. The Vicalloy tape specimen (impregnated in Bakelite) also has low magnetostriction (11×10^{-6}) and low hardness (R.C. 33). The Mishima composition of 23 percent nickel, 12 percent aluminum, and 65 percent iron, cooled from 1300°C at the rate of 5°C per second had the lowest magnetostriction of the alloys actually measured, a value of -1.0×10^{-6} with an accompanying coercive force of 356 oersteds. The magnetostriction of the other Mishima composition (29 percent nickel, 13 percent aluminum and 58 percent iron) is plotted for the sand-cast state, for which the saturation magnetostriction is -15×10^{-6} and the coercive force 560 oersteds.

In Fig. 14 one observes that for the carbon hardening alloys measured there is an approximately linear relation between magnetostriction and coercive force but for the carbon-free alloys measured there is no such relation.

The fourth and final phase of this work was the measurement of several alloys of special interest. Figure 15 shows a magnetostriction curve on the Ferroxcube specimen. The magnetostriction was less than 1×10^{-6} for all values of the field up to 600 oersteds. This result is in accordance with the theory of ferrites in which oxides are present in the proper proportion to obtain approximately zero magnetostriction.[15] Figure 15 also shows curves of magnetostriction obtained on tape specimens (0.002 in.×0.250 in.) of iron-cobalt alloys having 65, 70, and 75 percent cobalt. The tape was given 95 percent cold reduction, wound into a specimen form (2.0 in. I.D., 2.5 in. O.D.) and the turns consolidated by impregnating in Bakelite. Values of 120, 130, and 114×10^{-6} were obtained respectively. A value of 80×10^{-6} had previously been reported[16] on the alloy containing 65 percent cobalt. In our specimens, we have oriented the domains transversely by drastic cold work so that the maximum change in length is observed when the domains are oriented by the applied field. The value of 130×10^{-6} obtained on the 70 percent cobalt alloy is one of the highest values of magnetostriction reported in the literature.

CONCLUSIONS

The curves obtained on Alnico V show that its magnetostriction depends upon heat treatment, the values ranging from 22 to 35×10^{-6}. The highest value of 35×10^{-6} was for the cooling rate which gives optimum permanent magnet properties. If we include the magnetic field treatment, the magnetostrictive values of Alnico V vary from 2.5 to 43×10^{-6} or more than a tenfold variation. The low longitudinal value is due to the fact that most of the domains are oriented either parallel or antiparallel to the magnetization and a 180° reversal of the domains does not contribute to the change in length. The high transverse value is due to the fact that most of the domains are oriented at right angles to the magnetization and each domain contributes to a change in length. The behavior of magnetostrictive specimens of Alnico V which were treated in longitudinal and transverse fields was in accordance with domain theory, as has been pointed out by Hoselitz and McCaig.[17] They concluded from their measurements that the domain magnetization is along the easy crystallographic direction which makes the smallest angle with the axis of anisotropy. However, our data indicate that the domain magnetization is more closely aligned with the direction of the field applied during heat treatment (see Figs. 9 and 10).

The alloys in the immediate vicinity of Alnico V have yielded substantial but not quite as high values of magnetostriction as Alnico V. The alloy containing 6 percent aluminum had a maximum magnetostriction of 25.5×10^{-6}, the standard 8 percent aluminum alloy had a maximum magnetostriction of 35×10^{-6} while the 10 percent aluminum alloy had a maximum value of 19×10^{-6}.

As a result of the investigation of the magnetostriction of permanent magnets having various coercive forces it appears that for alloys hardened by carbon (tool steel, cobalt steels) there is a rough linear relationship between magnetostriction and coercive force for the alloys tested. When we include the carbon-free magnets in the comparison such as Alnico, Vicalloy, Remalloy, Cunico, and the iron-nickel-aluminum system, the linear relation no longer exists. For example Alnico V has over twice the coercive force of Honda steel but its magnetostriction is only 60 percent of that of Honda steel. Likewise the Remalloy and Vicalloy specimens measured had ap-

[15] J. L. Snoek, *New Developments in Magnetic Materials*, (Elsevier Publishing Company, Inc., Amsterdam, 1947).

[16] Y. Masiyama, Sci. Repts. Tohôku Imper. Univ. **21**, 394-410 (1932).

[17] Hoselitz and McCaig, Proc. Phys. Soc. **62B**, 163 (1949).

proximately the same coercive force but the magnetostriction of the Vicalloy was twice that of Remalloy. Cunico had a low magnetostriction and a low indentation hardness compared with its high coercive force of 445 oersteds. Finally many of the iron-nickel-aluminum alloys had a low magnetostriction combined with a high coercive force and for some of these alloys the magnetostriction actually became zero (Fig. 13).

On the basis of this experimental work one can say that the coercive force of some of the iron-nickel-aluminum alloys does not depend on the magnetostriction. Also, it may be concluded that in the other carbon-free alloys investigated, the magnetostriction is not the only factor operating in these systems to produce the high coercive force, especially in the case of Alnico V since this alloy is directly related to the iron-nickel-aluminum system. For the three carbon-hardening alloys investigated it appears that there may be a close relation between magnetostriction and coercive force.

Recent theories of coercive force have emphasized inhomogeneity of composition and fine particles as well as internal stress. The results obtained in this paper suggest that inhomogeneity of composition or the existence of fine particles is the controlling factor in the iron-nickel-aluminum system and that they may be a factor in the other carbon-free alloys investigated.

ACKNOWLEDGMENTS

The author is greatly indebted to Dr. R. M. Bozorth for discussions and suggestions given during the work, to Miss M. Goertz who assisted in making the measurements, and to Mr. D. H. Wenny who supervised the making of the alloys used.

Improved Ultra-Thin Sectioning of Tissue for Electron Microscopy*·†

JAMES HILLIER
RCA Laboratories, Princeton, New Jersey and Sloan-Kettering Institute for Cancer Research, New York, New York

AND

MARK E. GETTNER
Sloan-Kettering Institute for Cancer Research, New York, New York

(Received March 30, 1950)

The use of a reservoir of liquid to collect ribbons of thin sections as they are cut enables serial sections to be obtained and provides means for detecting variations in thickness. Reducing the effects of static friction and of external vibrations makes it possible to cut long series of sections 0.2μ thick and uniform to better than five percent. Some observations which may throw some light on the mechanics of the cutting process are reported. Mounting the sections without any treatment is shown to be superior to any technique in which the embedding material is removed or replaced. It is shown that the technique provides sections in which a structural resolution of 200A is possible. The remaining artifacts introduced by the sectioning and mounting are easily recognized deformations which do not influence the interpretation.

I. INTRODUCTION

ANY analysis of the requirements for the widespread application of the electron microscope in the fields of biology and medicine leads directly to the necessity of sectioning tissue routinely at thicknesses of the order of 0.2μ, preferably serially. Its acceptance by the cytologists, who will ultimately be the users, will depend entirely on the ability of the technique to provide resolution of structure and freedom from artifacts considerably beyond those obtained by the conventional techniques of histology and cytology. Pease and Baker[1] made a major contribution to this work when they demonstrated that it is feasible to cut sufficiently thin sections by relatively simple modifications of conventional microtomes and techniques. Viewed in retrospect, their subsequent work and that of the workers who followed their technique shows rather large scale artifacts. A simple check with a light microscope shows that these artifacts are of considerably greater magnitude than would be expected from the fixation and imbedding used. It must be assumed therefore, that the artifacts were introduced by the sectioning technique or the subsequent manipulation of the sections. The research reported in this paper involves a careful analysis of the Pease and Baker technique with the result that many of the sources of artifacts have been identified and removed.

The complete problem of examining tissues with the electron microscope involves: (1) the selection and fixation of the tissue, (2) embedding, (3) sectioning, and (4) mounting of sections. The first two steps include a number of variables about which no information can be

* This work was sponsored in part jointly by the ONR under Contract No. N6-ori-99 Task Order I and the AEC and in part by the Lillian Babbitt Hyde Foundation. The electron microscope used was supplied through the kindness of New York University.

† Based on a paper presented at the Washington D. C. meeting of the Electron Microscope Society of America, October, 1949 [J. App. Phys. **21**, 67A (1950)] supplemented by more recent work.

[1] D. C. Pease and R. F. Baker, Proc. Soc. Exper. Biol. and Med. **67**, 470–474 (1948).

obtained until the last two are developed to the point where routinely reproducible results are possible. Thus the work described in the following was all carried out on a set of specimen blocks obtained from osmium-perfused mouse liver. Actually the livers from only two mice were used and all the blocks from each liver were embedded simultaneously by the technique described in Pease and Baker's original work.[1]

II. SECTIONING: BLOCK ADVANCE

In spite of its multiplicity of moving parts, the Spencer rotary microtome (No. 820) has proved to be ideal for the cutting of thin sections. In this machine the knife is stationary and the block is moved vertically to make the cuts and advanced horizontally between cuts. In the mechanical design of the Spencer microtome, these two motions are intimately related and will be considered in the discussion concerning the advancing of the block. It was found that for very thin sections which require little cutting force the actual sectioning (that is, the action of the knife on the block) is not related to the advancing of the block except under some rather special conditions.

The reservoir for collecting and spreading the specimens serially, described by Gettner and Hillier[2,3] also proved very useful for testing the operation of the advance mechanism. It was found that for a given block, knife, cutting speed, and reservoir liquid the length of the floating section (in the direction of cutting) is a direct function of the thickness of the cut. The shortening of the sectioning is first appreciable at 0.3μ and increases rapidly with decreasing thickness, usually being greater than 50 percent for sections of 0.1μ thickness. Thus, the lengths of successive sections observed at a magnification of 27, as they appear, provides a sensitive criterion of the uniformity of thickness. By this means it has been possible to identify several sources of variations in the advance mechanism: (1) static friction, (2) vibrations of external origin, and (3) thermal expansion. In an unchanged Spencer microtome, operated by hand, these effects produce variations in thickness of the order of 0.2μ which while insignificant in normal sectioning are much too great for ultra-thin sectioning. *It is to be noted, from the discussion which follows, that modification of the method or magnitude of advance has no effect on these sources of variations (except possibly a slightly detrimental one).* Further refinements are necessary therefore, if sections of uniform thickness at the 0.1μ level are to be cut properly.

The advance mechanism of the Spencer microtome consists of a series of mechanical reductions.[4] At the top of each cutting cycle, a pawl actuated by the drive wheel engages a toothed ratchet wheel and advances it one tooth for the unit advance. Coaxial with, and rigidly attached to the wheel is an accurately threaded rod which engages a threaded nut. The nut in turn is rigidly attached to a pointer which rides against an inclined plane, an integral part of the specimen block holding and moving system. The forward motion of this last unit is guided by a set of horizontal slides which also support it. A spring keeps the inclined plane in contact with the pointer. The cutting is accomplished by moving the block and its support, including the spring, inclined plane, and horizontal slides, in a set of vertical slides. The cutting motion is sinusoidal and has an amplitude of 5 cm. Since the pointer is stationary, it traces a line on the inclined plane approximately vertical and approximately perpendicular to the inclination. If the normal to the inclined plane were in the horizontal plane, or, more precisely, perpendicular to the direction defined by the vertical slides, the true motion of the specimen block would be parallel to the vertical slides. Under this condition there would be no motion of the block support relative to the horizontal slides during the cutting movement. Actually, it is an important feature of the Spencer microtome that the normal to the inclined plane usually makes a small angle to the horizontal. This means that, as the block makes its vertical excursion of 5 cm, it also makes a horizontal excursion of several microns (300 in our case). Thus, there is relative motion in the horizontal slides some time before and during the actual cutting. By this simple means, the large varia-

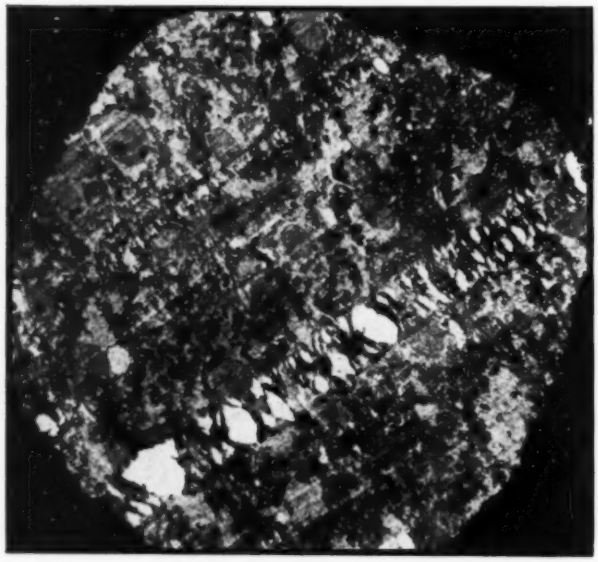


FIG. 1. Normal mouse liver cut at 0.2μ. Embedding was completely replaced with collodion. Micrographed with a double objective with limiting aperture set at 4×10^{-3} radian. This section demonstrates a uniform compression of approximately 40 percent which means that the actual specimen is more than 0.3μ thick. Destructive "knife marks" due to edge defects 8 and 2μ long are visible. All of the micrographs presented in this paper were with a standard RCA Model u.c. operating at 50 kv. Magnification approximately 900×.

[2] M. E. Gettner and J. Hillier, Science (submitted for publication).

[3] M. E. Gettner and J. Hillier, J. App. Phys. **21**, 68A (1950).

[4] O. W. Richards, *The Efficient Use and Proper Care of the Microtome* (American Optical Company, Buffalo, New York, 1949), second edition, p. 84.

tions in the advance of the block caused by static friction in the horizontal ways are almost completely eliminated. Some crude tests indicated that the force required to overcome the static friction in the horizontal slides was such as to produce a deformation of between 1 and 2μ in the inclined plane and block support. Thus if the horizontal excursion were eliminated and the advance set at 0.1μ, the specimen would still advance in 1 to 2μ steps and be cut only after every 10 to 20 cutting cycles. The horizontal excursion also means that any "play" in the vertical slides or any flexibility in any other associated parts is taken up by the same amount in each cycle.

It was found that for ultra-thin sections static friction in the horizontal slides still produced greater variations than were desirable. This was due to the fact that the relative motion in the horizontal slide stopped and reversed at the top and bottom of each cutting cycle allowing static friction to introduce variations which were not entirely removed by the time the block reached the knife. The addition of considerable extra weight in the modification of the inclined plane aggravated this situation. The variations in section thickness were then of the order of 0.05 to 0.1μ. In view of the reduced total advance in the modified system, it was possible to eliminate these variations by removing about 90 percent of the weight of the block support and modified inclined plane from the horizontal slides. To do this, a short coil spring of the proper tension was inserted between the rear point of attachment of the main spring and the lower brace for the inclined plane. While this was satisfactory, it was found more convenient to use a leaf spring mounted on the top of a new block which also serves to secure the back end of the main spring. The lifting force is applied to the block support by a bent rod which hooks under the lower brace of the inclined plane. The upper end of this rod is threaded and secured to the end of the leaf spring by a small nut. Thus the lifting force can be adjusted. The horizontal excursion of the block and the total advance are permitted by the fact that the leaf spring can twist and that neither end of the bent rod is rigidly attached.

The modification of the inclined plane used follows the principles described by Pease and Baker. There are, however, a number of specific points of difference. The modification consists of machined iron block which slips over the original inclined plane. It is held in place by eight pointed set screws which grip the original plane behind its top and bottom edges. There is close contact between the top and bottom edges of the original plane and the block but none in the central region. The new inclined plane is a selected piece of plate glass which is flat to within two fringes of sodium light. It is fastened to the iron block by embedding in Woods metal. Before embedding, the glass plate was adjusted to the proper position, by what amounts to three leveling screws threaded into the block.

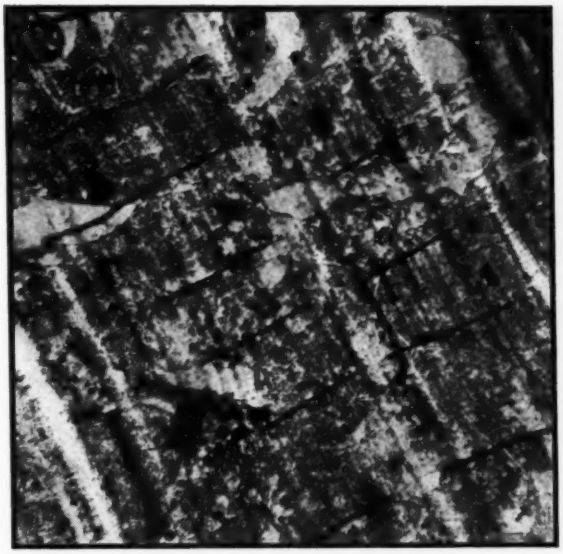


FIG. 2. Normal mouse liver cut at 0.2μ. Embedding was left intact. Micrographed with conventional objective but with focal length of approximately 8 mm and limiting aperture set at 3×10^{-2} radian. This shows a number of fine folds perpendicular to the direction of cutting caused, in this case, by the greater compression which occurred in the thinner strip which separates the two rows of folds. Note also that the "knife marks" are more evident than in Fig. 1. Magnification approximately 900×.

The unit of advance was set at 0.02μ. This low value was not chosen with any thought that sections of this thickness could be cut, but rather to overcome possible variations due to static friction in the slides that support and guide the pointer between the nut and the thread and in the end bearings of the thread and drive wheel. With this arrangement, sections of thickness 0.2μ which have been found most suitable require a rotation of the drive wheel and thread of 16.4° and a movement of the pointer of 20μ. Both of these movements are large enough to be accurately reproducible. This means also reduces the effects of inaccuracies in the threaded rod.

Variations of section thickness equal in magnitude to those put in by the effects of static friction are introduced by vibration of the microtome or of its support. The main trouble in this respect was traced to the irregularities of manual operation of the instrument. It was successfully overcome by adding a motor drive. A small 1/15 hp motor was used with a built-in reduction gear that gave 30 r.p.m. The motor was ballasted and shock mounted on sponge rubber separately from the microtome. In order to avoid transmission of vibrations through the coupling to the microtome and also to avoid difficulties of accurately aligning the drive shafts of the microtome and the motor, a special coupling was devised. The power is transmitted through a strip of Neoprene (approx. $\frac{1}{16}\times\frac{3}{8}\times1\frac{1}{2}$ in.) from the end of a 3-in. long bar clamped to the motor output shaft to a pin set in the microtome drive wheel, also 3 in. from the axis. The Neoprene seems to have the right degree of elasticity with the proper damping for the arrangement

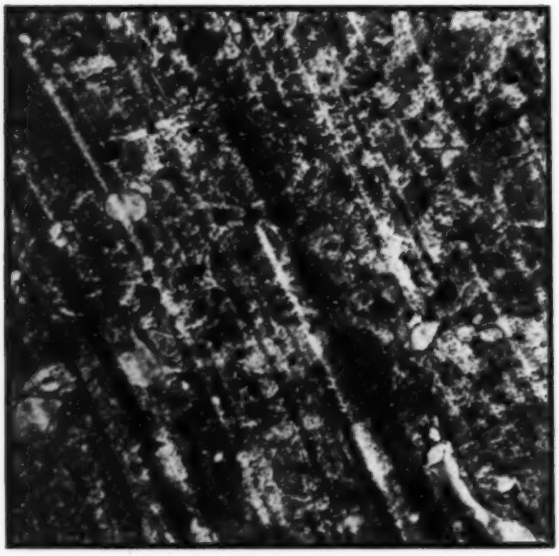


FIG. 3. All conditions identical with those of Fig. 2. Here a dense granule has been carried by the knife splitting the section until it was left on the section. Note that some of the denser "knife marks" appear to consist of a homogeneous material (paraffin?) which has been left on the specimen. Magnification approximately 900×.

used. Direct coupling by means of a wire loop or coupling through a coil spring under tension produced periodic variations in thickness.

A brass plate was added to the drive wheel of the microtome to counterbalance accurately the change in weight of the horizontal member caused by the addition of the modifications.

The instrument remains extremely sensitive to externally produced vibrations. In fact, resting a finger very lightly on any part of it will introduce variations in thickness amounting to 0.05 to 0.1μ. Building vibrations caused by nearby traffic also cause variations. Although it was not done in this work, it would be desirable to mount the instrument on a high inertia vibration-free support.

Thermal changes in the instrument are the only remaining major source of thickness variations. The time constant of these changes is so long, however, that with motor drive, the effect is more one of a disturbance of calibration. Thus, it cannot be detected directly by observation of the uniformity of compression as the sections appear on the liquid surface. It can be detected, however, by setting the advance to zero and observing the cutting through a low power microscope. A continuation of cutting or delay in returning to cutting when the advance is set to a finite value indicates the presence of thermal changes. More recently it has been found that an accurate estimate of the direction and amount of thermal drift can be obtained by the following procedure. The machine is allowed to cut a series of sections of desired thickness, say 0.2μ, then within one cycle the manual advance crank is turned backward one turn while the machine is allowed to continue operating. This represents a retraction of the block by 25 0.2μ sections. The knife is observed and the cutting cycles counted until the first new section is cut. This should happen on the 26th cycle. Any deviation will indicate the presence of thermal drifts. Furthermore, the amount and direction of the deviation will enable the true thickness of the cut to be calculated to within four percent in this example. More than one turn of the crank can be used if higher accuracy is desired. The method has the advantage that any disturbance introduced by touching the instrument is completely damped out by the time a new cut is made.

The most important source of thermal changes has been from the use of a focused light to illuminate the edge during observation. This practice has now been discontinued in favor of using a large fluorescent light placed some distance above the instrument. Ordinary changes in room temperature introduce little effect though it is essential that the instrument be placed in a location free from strong air currents.

A possible source of variations which has not yet been studied is in the thickness of the oil film which lubricates the vertical ways. Variations in the thickness of this film of 40A would be detectable under optimum conditions. Obviously such changes could occur. Furthermore, the "working in" of an oil film of different thickness could produce periodic variations. In practice, very little lubrication (Pike oil) has been used and the microtome is run at zero setting for several minutes before sections are cut. These are mainly precautionary measures and may prove unnecessary.

A final point which should be mentioned in connection with the function of the advance mechanism is in regard to the adjustment of the slides. This proved to be simpler than was expected because the cutting force required for the thin sections and very small blocks is very small. This force is never sufficient to lift the weight of the block supporting mechanism from the lower of the horizontal slides, nor to separate the rear of the vertical slides against the tension of the main spring. Thus, since a transference of the reactive force from one slide to the other never occurs in either set, there is no need to adjust the slides so that there would be no displacement of the block if such a transfer were to occur. In this work, the slides are adjusted so as to be very free running and are somewhat looser than would be considered good practice in ordinary microtomy.

III. SECTIONING: CUTTING

The interaction of the knife with the specimen which constitutes the cutting of sections is, as yet, a very poorly understood phenomenon. From this work it has become evident that defects in a knife-edge produce observable effects if they have dimensions of the same order as the thickness of the section cut. Thus it would appear that the cutting of ultra-thin sections will provide information regarding the nature of the edge at a level at least an order of magnitude below what was

previously possible and thus provide a very profitable approach to this fundamental problem.

It is apparent that the quality of the sections will be a complicated function of many variables, including in particular, the nature of the tissue, the fixation, the dehydration, the embedding, the sharpness and cleanliness of the edge, the bevel and clearance angles, the speed of cutting, the temperature, and the properties of the liquid in the reservoir. An attempt was made to reduce the number of variables by using only one specimen of fixed tissue cut into a number of pieces which were dehydrated and embedded simultaneously. Unfortunately, this leaves an uncomfortably large number of variables, at least two of which cannot be determined independently—namely, the sharpness and cleanliness of the edge. Nevertheless, some conclusions can be drawn from the observations already made.

For the work shown here, commercially sharpened knives were used throughout. It is recognized that such edges represent a compromise between quality and cost. More recently, hand-sharpened edges have been providing much superior results. The edges of the blades are examined at 400× by the scattered light technique described by Richards[2] and the facets by reflected light.

The fundamental defects in the sections obtained in this work consists of a uniform microscopic compression (Fig. 1), an accordian-like folding parallel to the cutting edge (Fig. 2) strips of partially destroyed tissue which run parallel to the direction of cutting (Fig. 1), and local distortions due to different cutting properties of different areas of the tissue. The last defect can vary in magnitude from the picking up by the blade of small hard granules which cut the sections until they are dropped (Fig. 3) to the folding of large areas. In passing, it should be mentioned that in material which is much harder than the embedding, there is a characteristic "chattering" phenomenon involving a rather uniform rearrangement of the tissue in 5 to 15μ steps.

The compression which appears as a microscopically uniform contraction in the direction of cutting is dependent on the nature of the tissue and the thickness of the section. It is definitely not dependent on the rake angle over the 6° range used in the tests. This is contrary to the experience of Baker and Pease.[5] This may be due to the presence of the liquid surface.

The accordion-like folding is peculiar to thin sections and does not appear in sections thicker than 0.3μ. It appears that the great flexibility of the sections permit them to fold whenever they encounter mechanical resistance beyond the cutting edge of the knife. As would be expected, the persistence of the folds is very sensitive to the cleanliness of the knife and to the surface tension and level of the liquid. It has been found necessary to adjust the surface tension and level of the liquid for many of the blocks. Observation of the sections at approximately 30× as they are cut provides a good criterion for these adjustments. It has been found that good sections appear suddenly and fully extended on the surface of the liquid. Moreover, they are transparent and nearly invisible. Poor sections appear to slide slowly on to the liquid from a folded condition near the edge of the blade and appear somewhat translucent.

A very special type of folding which cannot be removed is caused by a defect in the knife which cuts a thin track through the specimen without causing any destruction. The thinner line is more compressed with the result that the adjacent areas must fold to compensate for the change in length of the section (Fig. 2).

Large defects in the knife will simply split the section, but more often, a characteristic line of destruction is produced (Fig. 1). The latter occurs for nicks or other defects which are isolated and occupy more than 1μ of the length of the edge. Defects occupying a length of the edge approximately twice the thickness of the section appear to produce slight local deformations of the structure (Fig. 4).

Observations of long series of sections cut at one point of the knife indicate that the dulling is a step-by-step process involving an increase in the number of larger defects in the knife. A relationship has also been observed between the sharpness of the blade and the frequency with which it must be cleaned. A slow decrease in the lengths of successive sections and an immediate return to the original length upon cleaning is indicative of an edge that is in need of sharpening. Examination of sections produced by such an edge show a greatly in-

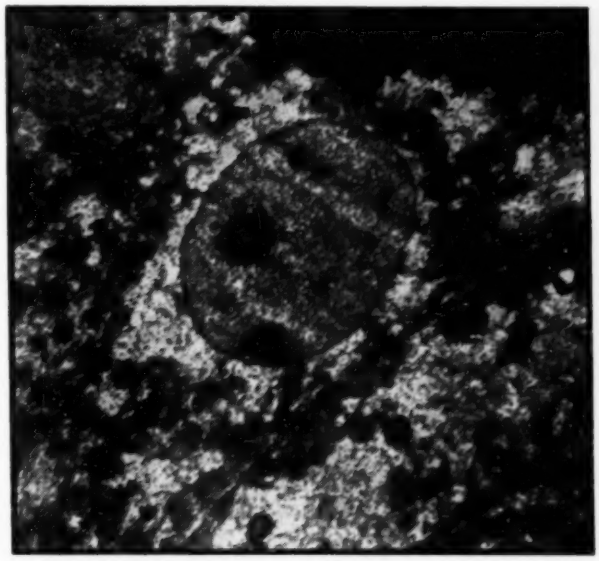


FIG. 4. A portion of the same section as shown in Fig. 1 taken at higher electronic magnification and with the same instrumental conditions. The nuclear membrane appears to be slightly deformed by a defect in the knife roughly estimated to be 0.5μ long. In comparing this with Fig. 6, note the thicknesses of the cell wall, the nuclear membrane, and the intracellular fibers. Note also that the granules in the cytoplasm appear hollow with dense walls. Magnification approximately 3,700×.

[5] R. F. Baker and D. C. Pease, J. App. Phys. **20**, 480 (1949).

creased number of lines of destruction which are also somewhat wider than obtained with a sharp edge.

Another interesting observation on the nature of the edge can be made by retracting the block a few section thicknesses and moving the blade, parallel to itself, to a new position. If the first cuts are observed, it is found that they are invariably narrow strips, all parallel to the direction of cutting, which grow wider on each successive cut until a complete section is obtained. This usually requires a total advance of the block of something more than a micron. This clearly indicates that the cutting edge is not a straight line but is corrugated with an amplitude of approximately 1μ.

IV. MOUNTING AND MICROGRAPHY OF SECTIONS

The handling of the specimens after they have been cut depends to a certain extent on their thickness, the nature of the tissue, and the problem under investigation. From our work, it appears that earlier workers have grossly underestimated the delicacy of the cut sections and that it is in this stage of the work that most of their artifacts have been introduced.

A technique described by Pease and Baker[3] involved mounting the section on a glass slide, after it had been spread on dioxane, and allowing it to dry. It was then washed with benzene to dissolve out the paraffin. This was followed with 0.2 percent collodion in amyl acetate which washed out the collodion embedding but left in a thin layer to support the specimen. The section was then floated off the slide on to clean water and mounted on a 200-mesh screen for examination in the microscope.

This technique was found to introduce artifacts in two ways. While the dioxane is an excellent means of spreading and flattening the section at the concentrations used, it nevertheless introduces considerable distortions in a way that is not completely understood. The use of the dilute collodion solution meant that the final film was much thinner than the section and thus, in the final stages of drying, surface forces are allowed to act on the very fragile structures of the section. This produced gross distortion of all but the coarsest structures (Fig. 5). The technique was then modified to omit the dioxane and to use a two percent solution of collodion leaving enough of the solution on the section to dry down to a film that was slightly thicker than the section. This eliminated the distortions due to surface forces, but greatly reduced the contrast obtainable and introduced difficulties in making the sections adhere to the specimen screens.

FIG. 5. Section cut at 0.2μ from which the embedding was entirely removed and replaced with collodion in amyl acetate of such concentration and amount as to produce a dried film of approximately 400A thickness. That there are large scale distortions is obvious from comparison with 2, 3, and 6. No objective aperture. Magnification approximately 900×.

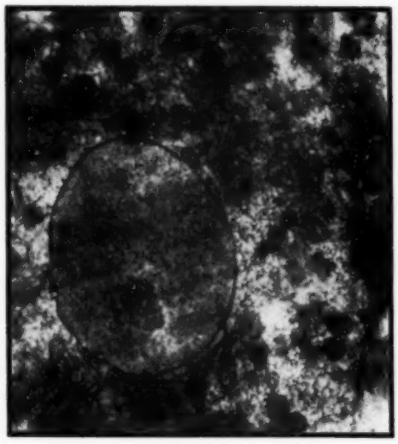


FIG. 6. A field selected for comparison with Fig. 4. The conditions of sectioning and micrography are the same with the exception that the embedding was not removed. Magnification approximately 3000×.

The contrast problem was overcome by raising the specimen to give an objective focal length of approximately 7 mm and by using a standard (50μ diam.) objective aperture resulting in an angular aperture of 3.5×10^{-3} radian.

This modified technique was finally abandoned when serial sections showed that merely dissolving the embedding material at any time after sectioning leaves many of the finer structures free to move. (Compare Figs. 4 and 6.)

The mounting technique now used involves no treatment of the section other than a mechanical mounting on a specimen screen (Figs. 6 and 7). A piece of 200-mesh copper screen (Elektromesh) 2×3 in. is etched in dilute nitric acid until it has at least 50 percent open area. A 300A thick collodion membrane is then mounted on this screen in the conventional way and dried. The sections

are mounted on suitable pieces of this screen by lifting them directly from the collecting reservoir. It is important that the prepared screen be lowered in to the liquid in a nearly vertical position but with the collodion membrane on the underside and that it be moved slowly through the liquid with film-side facing the direction of motion. After the sections are lifted from the liquid, the blank side of the screen is immediately blotted with hard dry filter paper and as much liquid as possible is drawn out from between the sections and the film.

The use of the extra film has successfully overcome the difficulty of anchoring the sections to the screen. In these sections there remains a problem of shrinkage under moderate electron bombardment. It can be avoided by raising the intensity quite slowly on each new field. Shrinkage of the section from around the screen openings can be avoided by placing the mount in the microscope with the sections facing the electron source, so that the whole area bombarded is seasoned uniformly.

The use of some dioxane (less than 50 percent) in the reservoir has been found to improve the flattening of the section in some cases, but requires that they be treated much more carefully under bombardment.

V. EXPERIMENTAL RESULTS AND DISCUSSION

Using the techniques described above, a large number of micrographs of the osmium-perfused liver have been obtained in which the structures have been consistent. The structure resolution is between 200 and 300A for the 0.2μ sections used. It is obvious that many new structures are being observed though their significance and even their reality are yet to be determined. Several artifacts introduced by the sectioning technique are still present. These include compression, folding, knife marks, shrinkage under bombardment, and occasional removal or shifting of material. However, all of these defects are quite obvious and do not interfere with the correct reading of the micrographs. It would be desirable, of course, to eliminate these as they detract from the esthetic value of the micrographs and can be annoying.

As indicated in the beginning, this investigation concerned only the factors involved in the cutting of uniformly thick sections and in the mounting for electron microscopic examination. The tissue, fixation, and embedding was maintained as constant as is possible in biological preparations. A type of tissue and a technique which had already been shown to be quite favorable for the problem were chosen. It is to be noted, however, that the refinements introduced are only slightly dependent on the nature of the preparation. This is not true of the cutting properties which can vary widely according to the nature of the tissue of the fixative, or of the embedding.

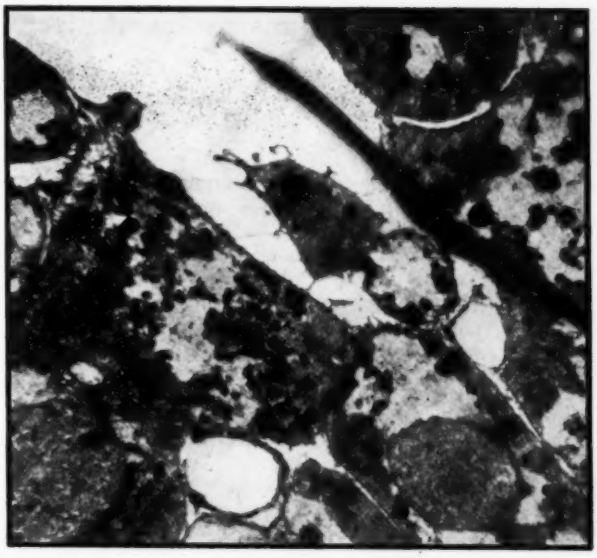


FIG. 7. A representative field of a 0.2μ section mounted with the embedding intact. Magnification approximately 1800×.

A number of other preparations have now been sectioned with varying degrees of success. However, while experience in this respect has not suggested any modifications in the techniques described above, it has emphasized the necessity of adjusting the embedding very carefully to the properties of the specimen and the technique of fixation and the need for sharper knives.

ACKNOWLEDGMENTS

The authors wish to thank Drs. J. J. Biesele and J. A. Jacquez for providing carefully prepared specimens and for their valuable discussions of the resulting micrographs. Thanks are also due Miss G. Cohn for her assistance in the embedding and mounting of the blocks.

Reflection and Transmission of Electromagnetic Waves by Thin Curved Shells*

JOSEPH B. KELLER
Institute for Mathematics and Mechanics, New York University, New York, New York
(Received March 31, 1950)

The scattering of an arbitrary electromagnetic field by a conducting or non-conducting obstacle is investigated. The differential equations and boundary conditions satisfied by the field are transformed into a pair of inhomogeneous linear integro-differential equations for E and H. For an obstacle which is a thin shell of constant thickness h, a formal procedure for obtaining a solution of these equations as power series in h is given. The lowest order term in this solution is the incident field. An explicit expression for the next term is found in the form of a surface integral. This integral is evaluated approximately by the method of stationary phase. The physical properties of the solution are examined in detail, and satisfactory agreement is found with many results previously obtained by other methods.

1. INTRODUCTION

WE wish to find the electromagnetic field which results when a given incident field strikes an obstacle—more precisely the field produced by a given source in the presence of an obstacle. For some special obstacles the problem has been solved exactly, and for arbitrary obstacles several approximate methods are available (geometrical optics,[1–3] Kirchhoff method,[3,4,5–7] Rayleigh scattering formula). Recently, Professor H. Primakoff and the author treated the corresponding acoustic problem for thin shell obstacles by a method which takes account of propagation within the obstacle and includes the results of the methods listed above.[3] In this paper the same method is applied to electromagnetic scattering by thin shells.

In Section 2 the problem is formulated and converted into an inhomogeneous integral equation. In Section 3 an approximate solution is obtained for an obstacle which is a thin shell. In Section 4 the integral representing the approximate solution is evaluated and in Section 5 the solution is examined and compared with other results. Many of the omitted details may be found elsewhere.[3,8]

2. DERIVATION OF THE INTEGRAL EQUATION

An obstacle (b) is immersed in an infinite medium (a), both being homogeneous and isotropic. A periodic source ($e^{-i\omega t}$) is located in (a). Outside the source the electric field $E(r)$ satisfies the reduced wave equation

$$(\nabla^2+k^2)E(r)=0. \tag{1}$$

Here $k^2=(\omega^2/c^2)\epsilon'\mu$; $\epsilon'=\epsilon+(4\pi i\sigma/\omega)$; $k\equiv k_a$ in (a); $k\equiv k_b$ in (b). (We shall write equations for E only; the corresponding equations for the magnetic field $H(r)$ are obtained by interchanging E with H and ϵ' with $-\mu$.) At the boundary between the two media the field satisfies the discontinuity conditions

$$E_t{}^a=E_t{}^b; \quad \epsilon_a'E_n{}^a=\epsilon_b'E_n{}^b. \tag{2}$$

Here n and t indicate normal and tangential components, and a, b indicate sides of the boundary, respectively. At infinity the scattered field $E-E_0$ satisfies a radiation condition. The source is characterized by $E_0(r)$, the field which would be produced in the absence of the obstacle. (We assume the obstacle does not affect the source.)

All these conditions and equations will be transformed into an integro-differential equation with the aid of a Green's function $G(r, r')$ which satisfies Eq. (1), with $k=k_a$ everywhere, with respect to the variable r'. At $r'=r$, G has a singularity of the form $1/(|r-r'|)$, and at infinity it satisfies a radiation condition. We apply Green's theorem to $G(r, r')$ and $E(r')-E_0(r')$, assuming $E-E_0$ has no singularities. We first integrate over the obstacle V bounded by S', then over the rest of space, and add, obtaining (r' is the integration variable)

$$E(r)-E_0(r)=\frac{1}{4\pi}\int_{S'}\left\{(E^a-E^b)\frac{\partial G}{\partial n}-G\left(\frac{\partial E^a}{\partial n}-\frac{\partial E^b}{\partial n}\right)\right\}ds+\frac{k_b{}^2-k_a{}^2}{4\pi}\int_V EG\,dv. \tag{3}$$

* The research reported in this document has been made possible through support and sponsorship extended in part by the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories, AMC, Contracts W28-099-ac-172 and with Washington Square College, New York University.

[1] H. J. Riblet and C. B. Barker, J. App. Phys. **19**, 63 (1948).

[2] J. B. Keller and H. B. Keller, J. Opt. Soc. Am. **40**, 48 (1950).

[3] H. Primakoff and J. B. Keller, J. Acous. Soc. Am. **19**, 820 (1947).

[4] J. B. Keller, "Reflection of electromagnetic waves," Special Report No. 172-1 to the Watson Laboratories; New York University, Washington Square College Mathematics Research Group, (December, 1946).

[5] Lord Rayleigh, *Theory of Sound, II* (Dover Publications, New York, 1945), pp. 118–126.

[6] Willis, "Reflection and refraction of ASDIC beams from various bodies and surfaces," British Admiralty Report (1943). (Integrals evaluated by Fresnel zone method.)

[7] "The scattering of radiation from rectangular planes, half-cylinders, hemispheres, and airplanes," report submitted to the Evans Signal Laboratory by the Moore School of Electrical Engineering, University of Pennsylvania (October, 1943). [Only linear terms retained in the phase (Frauenhofer approximation) which yielded easy integrals but some unreasonable results.]

[8] J. B. Keller, "Reflection and transmission of electromagnetic waves by thin curved shells," thesis, New York University (April, 1948).

To evaluate the first difference occurring in the integrand in Eq. (3), we use Eq. (2). We find (n is the unit normal)

$$E^a-E^b=[(\epsilon_b'/\epsilon_a')-1]E_n{}^b n. \tag{4}$$

The evaluation of the second difference is more difficult and requires the use of Eq. (2), Maxwell's equations and the use of curvilinear coordinates. The lines of curvature on the boundary surface S', designated as u_1=const., u_2=const. are employed. We finally obtain (see reference 8 for details)

$$\frac{\partial E_n{}^b}{\partial n}-\frac{\partial E_n{}^a}{\partial n}=\left(\frac{\epsilon_b'}{\epsilon_a'}-1\right)E_n{}^b\cdot 2G_m, \tag{5}$$

$$\frac{\partial E_1{}^b}{\partial n}-\frac{\partial E_1{}^a}{\partial n}=\frac{i\omega}{c}(\mu_b-\mu_a)H_2{}^b+\left(1-\frac{\epsilon_b'}{\epsilon_a'}\right)\frac{\partial E_n{}^b}{\partial u_1}, \tag{6}$$

$$\frac{\partial E_2{}^b}{\partial n}-\frac{\partial E_2{}^a}{\partial n}=-\frac{i\omega}{c}(\mu_b-\mu_a)H_1{}^b+\left(1-\frac{\epsilon_b'}{\epsilon_a'}\right)\frac{\partial E_n{}^b}{\partial u_2}. \tag{7}$$

Here u_1 and u_2 represent arc length along the lines of curvature; E_1 and E_2 represent components along these lines and G_m is the mean curvature of the surface S'. (A curvature is positive if the center of curvature and positive normal are on opposite sides of S'.)

We now use Eqs. (4)–(7) in Eq. (3) and obtain (i_1, i_2 are unit vectors along the lines of curvature)

$$E(r)=E_0(r)+\frac{1}{4\pi}\int_{S'}\left\{\left(\frac{\epsilon_b'}{\epsilon_a'}-1\right)E_n{}^b\left(\frac{\partial G}{\partial n}+2GG_m\right)n+G\left[\frac{i\omega}{c}(\mu_b-\mu_a)(i_1H_2{}^b-i_2H_1{}^b)+\left(1-\frac{\epsilon_b'}{\epsilon_a'}\right)\left(i_1\frac{\partial E_n{}^b}{\partial u_1}+i_2\frac{\partial E_n{}^b}{\partial u_2}\right)\right]\right\}dS+\frac{k_b{}^2-k_a{}^2}{4\pi}\int_V EG\,dV. \tag{8}$$

Equation (8) is the fundamental integro-differential equation for the field in the presence of an obstacle. The integrals represent the scattered field, and obviously vanish when the properties of the two media are the same.

3. SOLUTION FOR THIN SHELL OBSTACLES

Equation (8) may be simplified and solved if the obstacle is a thin shell of constant thickness h. Then S' consists of two parallel parts S_I and S_{II}, neglecting the edges. If r is exterior to the shell we may expand the integrands of Eq. (8) in Taylor series about points of S_I. Then Eq. (8) becomes approximately, for r in medium (a) (n is a unit normal to S_I pointing into the obstacle)

$$E(r)=E_0(r)+\frac{1}{4\pi}\int_{S_I}\sum_{m=0}^{\infty}\frac{h^{m+1}}{(m+1)!}\frac{\partial^m}{\partial n^m}\left[\frac{\partial}{\partial^n}\left\{\left(\frac{\epsilon_b'}{\epsilon_a'}-1\right)\left(\frac{\partial G}{\partial n}+2GG_m\right)E_n{}^b n+\frac{i\omega}{c}(\mu_b-\mu_a)G(i_1H_2{}^b-i_2H_1{}^b)+\left(1-\frac{\epsilon_b'}{\epsilon_a'}\right)G\left(i_1\frac{\partial E_n{}^b}{\partial u_1}+i_2\frac{\partial E_n{}^b}{\partial u_2}\right)\right\}+(k_b{}^2-k_a{}^2)GE^b\right]dS. \tag{9}$$

If the edges are taken into account, we obtain another similar line integral over the curve bounding S_I in Eq. (9).

To solve Eq. (9) we represent the field in medium (a) as a power series in h,

$$E(r)=\sum_{\nu=0}^{\infty}h^\nu E_{(\nu)}(r); \tag{10}$$

insert this into Eq. (9) and equate coefficients of like powers of h. It is to be understood that all quantities in Eq. (9) evaluated on side b are to be replaced by appropriate quantities on side a. We then find that $E_{(0)}=E_0$, and that all succeeding terms can be determined successively as surface integrals.

For $E_{(1)}(r)$ we obtain

$$E_{(1)}(r)=\frac{1}{4\pi}\int_{S_I}\left\{n\left[\left(\frac{\epsilon_b'}{\epsilon_a'}-1\right)\left(\frac{\partial^2 G}{\partial n^2}+2G_m\frac{\partial G}{\partial n}\right)E_{n(0)}{}^b+\left(\frac{\epsilon_b'}{\epsilon_a'}-1\right)\left(\frac{\partial G}{\partial n}+2GG_m\right)\frac{\partial E_{n(0)}{}^b}{\partial n}+(k_b{}^2-k_a{}^2)GE_{n(0)}{}^b\right]+i_1\left[\frac{i\omega}{c}(\mu_b-\mu_a)\left(\frac{\partial G}{\partial n}H_{2(0)}{}^b+G\frac{\partial H_{2(0)}{}^b}{\partial n}\right)+\left(1-\frac{\epsilon_b'}{\epsilon_a'}\right)\left(\frac{\partial G}{\partial n}\frac{\partial E_{n(0)}{}^b}{\partial u_1}+G\frac{\partial^2 E_{n(0)}{}^b}{\partial n\partial u_1}\right)+(k_b{}^2-k_a{}^2)GE_{1(0)}{}^b\right]+i_2\left[-\frac{i\omega}{c}(\mu_b-\mu_a)\left(\frac{\partial G}{\partial n}H_{1(0)}{}^b+G\frac{\partial H_{1(0)}{}^b}{\partial n}\right)+\left(1-\frac{\epsilon_b'}{\epsilon_a'}\right)\left(\frac{\partial G}{\partial n}\frac{\partial E_{n(0)}{}^b}{\partial u_2}+G\frac{\partial^2 E_{n(0)}{}^b}{\partial n\partial u_2}\right)+(k_b{}^2-k_a{}^2)GE_{2(0)}{}^b\right]\right\}dS. \tag{11}$$

The integrand involves the field and its derivatives on the b side of S_I, and these must be replaced, using

Eqs. (4)–(7), by corresponding quantities on the a side. Equation (11) then becomes

$$E_{(1)}(r)=\frac{1}{4\pi}\int_{S_I}\Big\{n\Big[E_{n(0)^a}\Big(1-\frac{\epsilon_a'}{\epsilon_b'}\Big)\Big(\frac{\partial^2 G}{\partial n^2}+2G_m\frac{\partial G}{\partial n}\Big)+\Big(\frac{\epsilon_b'}{\epsilon_a'}-1\Big)\Big(\frac{\partial G}{\partial n}+2GG_m\Big)\Big(\frac{\partial E_{n(0)^a}}{\partial n}+2G_m E_{n(0)^a}$$
$$\times\Big[1-\frac{\epsilon_a'}{\epsilon_b'}\Big]\Big)+(k_b^2-k_a^2)\frac{\epsilon_a'}{\epsilon_b'}GE_{n(0)^a}\Big]+i_1\Big[\frac{i\omega}{c}(\mu_b-\mu_a)\Big(\frac{\partial G}{\partial n}H_{2(0)^a}+G\frac{\partial H_{2(0)^a}}{\partial n}$$
$$+\frac{i\omega}{c}G[\epsilon_b'-\epsilon_a']E_{1(0)^a}+G\Big[\frac{\mu_a}{\mu_b}-1\Big]\frac{\partial H_{n(0)^a}}{\partial u_2}+\Big(\frac{\epsilon_a'}{\epsilon_b'}-1\Big)\frac{\partial G}{\partial n}\frac{\partial E_{n(0)^a}}{\partial u_1}+G\Big(1-\frac{\epsilon_b'}{\epsilon_a'}\Big)$$
$$\times\Big(\frac{\partial^2 E_{n(0)^a}}{\partial u_1\partial n}+2G_m\Big[1-\frac{\epsilon_a'}{\epsilon_b'}\Big]\frac{\partial E_{n(0)^a}}{\partial u_1}\Big)+(k_b^2-k_a^2)GE_{1(0)^a}\Big]+i_2\Big[-\frac{i\omega}{c}(\mu_b-\mu_a)\Big(\frac{\partial G}{\partial n}H_{1(0)^a}$$
$$+G\frac{\partial H_{1(0)^a}}{\partial n}-\frac{i\omega}{c}G[\epsilon_b'-\epsilon_a']E_{2(0)^a}+G\Big[\frac{\mu_a}{\mu_b}-1\Big]\frac{\partial H_{n(0)^a}}{\partial u_1}\Big)+\Big(\frac{\epsilon_a'}{\epsilon_b'}-1\Big)\frac{\partial G}{\partial n}\frac{\partial E_{n(0)^a}}{\partial u_2}+G\Big(1-\frac{\epsilon_b'}{\epsilon_a'}\Big)$$
$$\times\Big(\frac{\partial^2 E_{n(0)^a}}{\partial u_2\partial n}+2G_m\Big[1-\frac{\epsilon_a'}{\epsilon_b'}\Big]\frac{\partial E_{n(0)^a}}{\partial u_2}\Big)+(k_b^2-k_a^2)GE_{2(0)^a}\Big]\Big\}dS. \quad (12)$$

Equation (12) represents the scattered field to first order in h in medium (a). Only the integral over the edges has been neglected in obtaining this result. In the same way, the higher order terms in the solution may be obtained.

4. EVALUATION OF $E_{(1)}(r)$

To evaluate the integral for $E_{(1)}(r)$ in Eq. (12) we select for $G(r, r')$ the Green's function for infinite space:

$$G(r, r')=\frac{\exp(ik_a|r-r'|)}{|r-r'|}. \quad (13)$$

We assume that the incident field is given by

$$E_0(r)=\mathcal{E}(r)\exp(ik_a|r-r_0|), \quad (14)$$

where $\mathcal{E}(r)$ is slowly varying compared to the exponential factor, in the vicinity of the obstacle. This is exactly true for a plane wave ($\mathcal{E}=$const.) and for other sources it becomes more accurate as the distance from the source increases. The position r_0 is an equivalent center of the source. We now insert Eqs. (13) and (14) into Eq. (12), and consider the variations of the exponential factors only. Before writing the resulting equation, it is convenient to introduce the following angles (n is normal to S_I into obstacle):

$$\gamma=\text{angle between } r'-r_0 \text{ and } n;\ \text{thus } \cos\gamma=\frac{\partial|r'-r_0|}{\partial n}, \qquad \beta=\quad r'-r_0 \text{ and } i_1;\quad \cos\beta=\frac{\partial|r'-r_0|}{\partial u_1},$$
$$\delta=\quad r'-r_0 \text{ and } i_2;\quad \cos\delta=\frac{\partial|r'-r_3|}{\partial u_2}, \qquad \alpha=\quad r'-r \text{ and } n;\quad \cos\alpha=\frac{\partial|r'-r|}{\partial n}. \quad (15)$$

Then Eq. (12) becomes

$$E_{(1)}(r)\approx\frac{1}{4\pi}\int_{S_I}\Big\{n\mathcal{E}_n\Big[\Big(1-\frac{\epsilon_a'}{\epsilon_b'}\Big)(-k_a^2\cos^2\alpha+2ik_aG_m\cos\alpha)+\Big(\frac{\epsilon_b'}{\epsilon_a'}-1\Big)(ik_a\cos\alpha+2G_m)$$
$$\times\Big(ik_a\cos\gamma+2G_m\Big[1-\frac{\epsilon_a'}{\epsilon_b'}\Big]\Big)+(k_b^2-k_a^2)\frac{\epsilon_a'}{\epsilon_b'}\Big]+i_1\Big[\frac{i\omega}{c}(\mu_b-\mu_a)\Big(ik_a\mathcal{H}_2\cos\alpha+ik_a\mathcal{H}_2\cos\gamma$$
$$+\frac{i\omega}{c}[\epsilon_b'-\epsilon_a']\mathcal{E}_1+ik_a\mathcal{H}_n\Big(\frac{\mu_a}{\mu_b}-1\Big)\cos\delta\Big)-\Big(\frac{\epsilon_a'}{\epsilon_b'}-1\Big)k_a^2\mathcal{E}_n\cos\alpha\cos\beta+\Big(1-\frac{\epsilon_b'}{\epsilon_a'}\Big)$$
$$\times\Big(-k_a^2\mathcal{E}_n\cos\alpha\cos\beta+2ik_aG_m\mathcal{E}_n\Big(1-\frac{\epsilon_a'}{\epsilon_b'}\Big)\cos\beta\Big)+(k_b^2-k_a^2)\mathcal{E}_1\Big]$$
$$+i_2\Big[-\frac{i\omega}{c}(\mu_b-\mu_a)\Big(ik_a\mathcal{H}_1\cos\alpha+ik_a\mathcal{H}_1\cos\gamma-\frac{i\omega}{c}[\epsilon_b'-\epsilon_a']\mathcal{E}_2+ik_a\mathcal{H}_n\Big[\frac{\mu_a}{\mu_b}-1\Big]\cos\beta\Big)$$
$$-\Big(\frac{\epsilon_a'}{\epsilon_b'}-1\Big)k_a^2\mathcal{E}_n\cos\alpha\cos\delta+\Big(1-\frac{\epsilon_b'}{\epsilon_a'}\Big)\Big(-k_a^2\mathcal{E}_n\cos\gamma\cos\delta+2ik_aG_m\mathcal{E}_n\Big[1-\frac{\epsilon_a'}{\epsilon_b'}\Big]\cos\delta\Big)$$
$$+(k_b^2-k_a^2)\mathcal{E}_2\Big]\Big\}\frac{\exp[ik_a(|r-r'|+|r'-r_0|)]}{|r-r'|}dS. \quad (16)$$

The integral in Eq. (16) can be evaluated approximately by the method of stationary phase. The result is (the factors P and $\mathcal{G}$ are given in Eqs. (19) and (18), respectively):

$$hE_{(1)}(r)=P\cdot\mathcal{G}\cdot\left[\frac{k_a h}{2}\right]\Big[n\mathcal{E}_n\Big\{\Big(1-\frac{\epsilon_a'}{\epsilon_b'}\Big)\Big(\frac{2iG_m}{k_a}-\cos\alpha\Big)+\Big(\frac{\epsilon_b'}{\epsilon_a'}-1\Big)\Big(i+\frac{2G_m}{k_a\cos\alpha}\Big)\Big(i\cos\gamma+\frac{2G_m}{k_a}\Big[1-\frac{\epsilon_a'}{\epsilon_b'}\Big]\Big)$$
$$+\Big(\frac{k_b^2}{k_a^2}-1\Big)\frac{\epsilon_a'}{\epsilon_b'\cos\alpha}\Big\}+i_1\Big\{\frac{i\omega}{k_a c}(\mu_b-\mu_a)\Big(i\mathfrak{IC}_2\Big[1+\frac{\cos\gamma}{\cos\alpha}\Big]+\frac{i\omega}{k_a c\cos\alpha}[\epsilon_b'-\epsilon_a']\mathcal{E}_1$$
$$+i\frac{\cos\delta}{\cos\alpha}\mathfrak{IC}_n\Big[\frac{\mu_a}{\mu_b}-1\Big]\Big)-\Big(\frac{\epsilon_a'}{\epsilon_b'}-1\Big)\cos\beta\mathcal{E}_n+\Big(1-\frac{\epsilon_b'}{\epsilon_a'}\Big)-\frac{\cos\gamma}{\cos\alpha}+\frac{2iG_m}{k_a\cos\alpha}\Big[1-\frac{\epsilon_a'}{\epsilon_b'}\Big]\mathcal{E}_n\cos\beta\Big)$$
$$+\Big(\frac{k_b^2}{k_a^2}-1\Big)\frac{\mathcal{E}_1}{\cos\alpha}\Big\}+i_2\Big\{-\frac{i\omega}{k_a c}(\mu_b-\mu_a)\Big(i\mathfrak{IC}_1\Big[1+\frac{\cos\gamma}{\cos\alpha}\Big]-\frac{i\omega}{k_a c\cos\alpha}[\epsilon_b'-\epsilon_a']\mathcal{E}_2+i\frac{\cos\beta}{\cos\alpha}\mathfrak{IC}_n\Big[\frac{\mu_a}{\mu_b}-1\Big]\Big)$$
$$-\Big(\frac{\epsilon_a'}{\epsilon_b'}-1\Big)\cos\delta\mathcal{E}_n+\Big(1-\frac{\epsilon_b'}{\epsilon_a'}\Big)\Big(-\frac{\cos\gamma}{\cos\alpha}+\frac{2iG_m}{k_a\cos\alpha}\Big[1-\frac{\epsilon_a'}{\epsilon_b'}\Big]\Big)\mathcal{E}_n\cos\delta+\Big(\frac{k_b^2}{k_a^2}-1\Big)\frac{\mathcal{E}_2}{\cos\alpha}\Big\}\Big]. \quad (17)$$

5. PROPERTIES OF THE SOLUTION

In Eq. (17) the first-order scattered (reflected, transmitted, or diffracted) field has been written as a product of three factors, each of which has special physical significance. It is interesting to observe that the solution of the corresponding acoustic problem has a similar form, the first two factors being identical with those found here.[3] We shall now discuss each of the factors in turn.

Geometrical Factor

The factor $\mathcal{G}$ in Eq. (17) depends upon geometrical quantities only and is given by

$$\mathcal{G}=\Big\{1+D_2\Big(\frac{1}{D_1}\Big[\frac{2-\sin^2\gamma}{\cos^2\alpha}-\tan^2\alpha+2\tan^2\alpha\sin^2\alpha\sin^2(\theta-\varphi)\Big]-(\cos\alpha+\cos\gamma)[2G_m+G_{||}(\theta)\tan^2\alpha]\Big)$$
$$+\Big(\frac{\cos\gamma}{\cos\alpha}\frac{D_2}{D_1}\Big)^2+D_2^2G_g\Big(1+\frac{\cos\gamma}{\cos\alpha}\Big)^2-\frac{D_2}{D_1}\frac{\cos\alpha+\cos\gamma}{\cos^2\alpha}(2G_m\cos^2\gamma+G_{||}(\varphi)\sin^2\gamma)\Big\}^{-\frac{1}{2}}\frac{\cos\alpha}{|\cos\alpha|}. \quad (18)$$

The various symbols in Eq. (18) are defined as follows, at the point on S_I where the phase in Eq. (16) is stationary: θ = angle between u_1 and the plane containing the normal and r_1, φ = angle between u_1 and the plane containing the normal and r_2, G_m = mean curvature, G_g = Gaussian curvature, $G_{||}(\theta)$ = curvature of the normal section of surface S determined by the plane making angle θ with u_1, $D_1=|r'-r_0|$, $D_2=|r'-r|$.

We say that this factor accounts for the geometrical spreading or focusing of the scattered field because it also occurs in the purely geometrical optical solution of the reflection problem. In fact, the ratio of the field amplitude reflected from an interface to the incident amplitude has been computed by geometrical optics[2,3] and found to be precisely the above quantity $\mathcal{G}$ in directions of specular reflection ($\alpha=\gamma$, $\theta=\varphi$) and zero in other directions (except for a reflectivity factor.) This factor has been investigated in references 2 and 3 and shown to yield the positions of caustics, foci, point images, focal length, and the mirror law. $\mathcal{G}$ becomes infinite on caustics and at foci, but a more careful evaluation of the integral in Eq. (16) at such places, employing Airy integrals instead of Fresnel integrals, yields finite but large values of the field.

Phase Factor

The factor P in Eq. (17) depends upon the incident wave-length and the obstacle dimensions, as well as the distances from source and observation point to the surface. It is given by

$$P=\frac{e^{iM'}}{2}[F_\pm(\xi_0'')-F_\pm(\xi_1'')][F_\pm(\eta_0'')-F_\pm(\eta_1'')]. \quad (19)$$

The symbols in Eq. (19) are defined as follows (see reference 3 for further details):

$F_\pm(x)=\int_0^x \exp[\pm i(\pi/2)u^2]du$ = Fresnel integral; plus or minus sign chosen in accordance with the rule given in reference 3.

$$M'=k_a(D_1+D_2)+\cdots.$$

P oscillates as the wave-length varies, with its absolute value ranging between zero and about two. It generally has a maximum in the direction of specular reflection, and approaches one in this direction, and zero in all other directions, as the wave-length approaches

zero, thus yielding the optical "law of reflection." In addition it yields a 180° phase change on crossing a point image or focus. Reflection from an interface has been treated by the Kirchhoff method[3,4] and the reflected field found to equal the incident field multiplied by $\mathcal{G}P$.

The arguments of the Fresnel integrals in Eq. (19) are given by

$$\xi_0''=\frac{\sqrt{2}\xi_0'}{\left(\frac{\pi}{|A'|}\right)^{\frac{1}{2}}},\quad \eta_0''=\frac{\sqrt{2}\eta_0'}{\left(\frac{\pi}{|C'|}\right)^{\frac{1}{2}}} \tag{20}$$

with similar expressions for ξ_1'', η_1''. The quantities $\xi_0'\xi_1'\eta_0'\eta_1'$ determine the boundaries of a rectangle equivalent to the "illuminated" part of the surface, referred to the point of specular reflection as origin. The denominators in Eq. (20) are just the half-dimensions of the first Fresnel zone on the surface around the specular reflection point. If the dimensions of the surface are large compared to those of the first Fresnel zone, the Fresnel integrals have their asymptotic values and the phase factor has absolute value one (provided medium a has zero conductivity, i.e., k_a is real). This is usually the case unless the dimensions of the surface are small compared to a wave-length or it is cylindrical or plane.

When the surface is small compared to the first Fresnel zone the arguments of the Fresnel integrals are small, and these integrals may be approximated by the first terms in their Taylor expansions, which yield $F_{\pm}(x)\approx x$. Then

$$P=e^{iM'}\frac{(\xi_0'-\xi_1')(\eta_0'-\eta_1')}{\pi/(|A'C'|)^{\frac{1}{2}}}. \tag{21}$$

From Eq. (21) we see that, apart from the exponential factor, P is equal to the ratio of the area of the obstacle surface to the area of the first Fresnel zone.

Since both A' and C' have the factor k_a, which is equal to $2\pi/\lambda_a$, where λ_a is the wave-length in medium (a), if the conductivity of medium (a) is zero, P is inversely proportional to λ_a. The reflectivity factor also has the factor $k_a h$, and hence one sees that the scattered field is proportional to the obstacle volume (=area$\times h$) and inversely proportional to λ_a^2. This is the Rayleigh scattering law for electromagnetic waves. In particular, it is valid if $\lambda_a \gg$ obstacle dimensions, since then the first Fresnel zone is larger than the obstacle.

Reflectivity Factor

The scattered field is expressed in Eq. (17) in terms of components in the n, i_1, i_2 directions, where n, i_1, i_2 are mutually orthogonal unit vectors at the point of stationary phase on the surface S. Each of these components is the product of the geometrical and phase factors and a third factor which is called the reflectivity factor. Since this factor is not the same for all components, it is appropriate to speak of three reflectivity factors, one for each component. The factors are the coefficients of n, i_1, and i_2 in Eq. (17).

The reflectivity factors are linear and homogeneous in the components of the incident field. They depend upon the physical properties of both media, the thickness and curvature of the shell, and the angles of incidence and reflection. Thus it can be seen that these factors take account of the propagation within the shell. From an examination of the form of these factors it is seen that they are proportional to the amplitude of the incident field, and increase in absolute value as $|\epsilon_a'-\epsilon_b'|$ and/or $|\mu_a-\mu_b|$ increase. Moreover, they are proportional to $k_a h$ which, in the case when medium (a) has zero conductivity, is $2\pi h/\lambda_a$. Thus the scattered field increases with the ratio of the thickness to the wave-length. This inverse dependence upon λ_a, in conjunction with the phase factor, leads to the electromagnetic analog of the Rayleigh scattering law, as has been pointed out above. When reflection and diffraction problems are treated by the Kirchhoff method no reflectivity factor is obtained, and hence the Rayleigh scattering law cannot be deduced.

If the reflectivity factors are divided respectively by the corresponding field components, they agree with the exact coefficients[9] for the reflection of plane waves from a thin, infinite, plane plate, except for additional terms proportional to the mean curvature of the surface. The polarization effects are therefore practically the same as for the plane case above. If Eq. (17) is specialized to the case of plane waves incident on a plane shell, the geometrical factor and the phase factor both become unity, and the reflected and transmitted fields agree exactly with those obtained by Luneberg[9] when his results are applied to a thin plate.

The presence of the quantity $\cos\alpha$ in the denominator of some terms of the reflectivity factors appears to indicate that the scattered field becomes infinite as $\alpha\rightarrow\pi/2$. In this case, however, the method used to evaluate the integral is not valid (see above).

CONCLUSION

The solution seems to yield a rather satisfactory quantitative as well as qualitative understanding of reflection and transmission by thin shells, as well as agreement with various special results previously obtained by other methods. For thick shells additional terms in the series solution would probably be needed. Since the calculations required would be quite complicated, an alternative method which was useful in acoustics seems preferable. We merely assume that the reflected field still contains three factors, the geometrical and phase factors being the same as those given here and the reflectivity factor being that obtained for a thick

[9] R. K. Luneberg, "The propagation of electromagnetic plane waves in plane parallel layers"; Research Report No. 172-3 to the Watson Laboratories; New York University, Washington Square College Mathematics Research Group (June, 1947).

plate. The physical interpretation of the various factors makes this procedure at least plausible. Calculations based on these assumptions seemed to agree with experimental results in acoustics.

A further test of the result of this paper is provided by an exact solution for the reflection and transmission of spherical waves by a spherical shell with a dipole source at its center.[10] This solution was specialized to the case of a thin shell and compared to the result of the present paper when applied to this case. The agreement between the two solutions was satisfactory for all components except E_r, the radial component of electric field, and except near the source (see reference 10 for further details).

[10] J. B. Keller and H. B. Keller, J. App. Phys. **20**, 393–396 1949).

ACKNOWLEDGMENT

I wish to express my appreciation to Professor Henry Primakoff for his excellent instruction, advice and direction in the solution of the corresponding acoustic problem.

A Technique for Taking Motion Pictures of Electron Microscope Images*

LUTHER E. PREUSS AND JOHN H. L. WATSON
Edsel B. Ford Institute for Medical Research, Detroit 2, Michigan
(Received April 12, 1950)

Lately, for the first time, a technique for taking motion pictures of electron microscope images has been developed which is proving useful in the study of some of the effects of electrons upon matter and particularly upon colloidal crystals. Resolution of 150 to 300A is achieved and is suitable for a host of practical electron microscopical problems. Pictures are taken from outside the instrument, utilizing the radiation emitted by the fluorescent screen. Some instrument adaptation is required for adequate illumination, and precautions must be exercised against toxic x-ray dosages.

THE feasibility of taking motion pictures of electron microscope images has been considered on occasion, but a lack of informative, moving subjects made the project impractical. Lately, the technique has been developed successfully in these laboratories for studying the effects of electrons, and indirectly of heat upon colloidal crystals. The effects of electrons upon a variety of other crystalline and non-crystalline materials are also being studied in the same way, and many specimen changes are observed in a much more positive manner than heretofore. Some reactions which escape detection entirely or which occur too rapidly for satisfactory observation to be made at all by the orthodox procedures are registered on motion picture films. Pronounced reactions have been recorded in one frame intervals in as short a time as 0.03 sec.

The motion pictures are taken by photographing the image on the fluorescent screen with the camera set up outside the microscope. Figure 1 shows the relative positions of camera and microscope with the final viewing screen tilted so as to have its plane normal to the optical axis of the camera and to give a maximum field of view. The angle is about 30 degrees. Some distortion is introduced by the tilt, and this has to be kept in mind when interpreting the images, but it is not found to be bothersome. The camera is a 16-mm Cine Kodak Special with a 4-in., $f/2.7$ telephoto lens mounted upon a conventional tripod. This tripod is a source of error due to vibration and a more rigid camera support is recommended. Practical operating speeds are from 8 to 32 frames per second, depending upon screen intensity. Above 32 frames per second the contrast deteriorates. Super XX negative panchromatic safety film has been found to produce the best contrast and Super XX and Kinolux positive emulsions are used with moderate success. The complete processing time with *D*-19, exclusive of drying, is at least two hours per 100-ft. roll of 16-mm negative. To reduce the amount of manual work required in the processing and to shorten the time, a motor-driven tank has been built, Fig. 2.

FIG. 1. Experimental arrangement of camera and microscope, showing angle of final viewing screen.

The intensity of the light emitted from the final viewing screen has to be of a relatively high order. A self-biasing electron gun[1] is the first requirement, both for sufficient intensity and for noticeable effects of bombardment upon samples. It is necessary to replace the fluorescent materials on the screens periodically because they become contaminated and lose their brilliance. Since much of the work has to be done in the lowest portion of the linear section of the Hurter-Driffield curve, the image brightness has to be increased to its maximum without provoking premature or undesirable specimen changes. In general, although the contrast would be desirable, no objective aperture can be used except with materials which react very slowly under the beam. Objective apertures were used in taking moving micrographs of colloidal crystals[2] of β-FeOOH under bombardment, but none were usable with NaCl or WO_3 crystals where the desired effects occurred more rapidly.

In order to increase the possible intensity still further, the diameter of the condenser aperture is opened from 25 to 50 mils. This, plus the tilting of the final viewing

* Presented in part before the annual meeting of the Electron Microscope Society of America, Washington, D. C. (1949).

[1] James Hillier and R. F. Baker, J. App. Phys. **17**, 12 (1946).
[2] John H. L. Watson, J. App. Phys. **19**, 713 (1948).

screen, results in the production of a considerable amount of x-radiation and makes it imperative that the instrument be surveyed. Some remarkably high dosage rates were measured for the experimental conditions. Figure 3 shows rough isodose curves at the position of the final viewing screen.

The electron microscope, which has inherent in it both a source of heat from electron bombardment in vacuum and a means of photography from the same radiation, offers a convenient means of conducting studies of solid-liquid transformations. The application of the moving picture technique provides a method for investigating the crystalline properties of such transformations by observation of Bragg electron reflections in specimens under bombardment. In normal electron microscopy Bragg reflections are seen to best advantage in dark field,[3] but in bright or dark field are most easily and definitely detectable when caused to move by variations of objective lens current. When the object itself is crystallizing or melting, or is in some other unstable condition, the images of the reflections move on the screen and are detectable without an objective current variation. They may be recorded effectively, therefore, by the motion picture technique. From direct observation, the bombarded samples are seen to emit a profusion of such reflections which originate from what appear from characteristics of shape alone to be melted areas as well as from those regions which are judged to be composed of well-formed crystals. Presumably, some of the ostensibly melted areas must often retain some oriented forms.

The motion picture technique for electron microscopes should also be useful for studying samples which are under the influence of factors other than electron bombardment in the microscope. For example, observations could be recorded to advantage upon film, if heat effects were studied, where the heating was done by a direct furnace source rather than by the beam. Ideal objects for study would also be provided during the deposition of thin metal films within the vacuum of the electron microscope, as reported[4] by Sennett of the Toronto group. The method is promising for a detailed study of the effects of electron bombardment upon bacteriological and biological specimens, and whenever a method is

[3] C. E. Hall, J. App. Phys. **19**, 198 (1948).

[4] R. S. Sennett and T. A. McLauchlan, University of Toronto, unpublished research, presented before the annual meeting of the Electron Microscope Society, Washington, D. C. (1949).

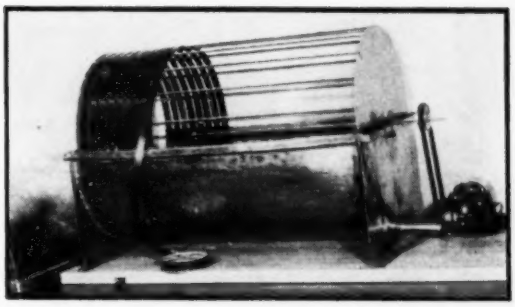


FIG. 2. Motor-driven, processing tank.

devised for studying live or wet specimens satisfactorily by electron microscopy the advantages of recording the observations on movie film will be obvious.

There are a number of other incidental advantages to be gained from taking motion pictures of electron microscope images. Non-electron microscopists, who are not as familiar with the appearance of images in the microscope but who are collaborating on a problem, find

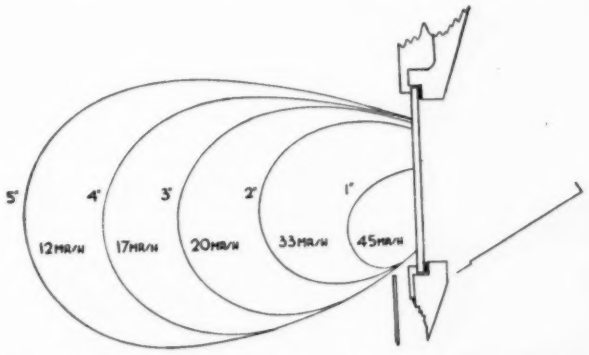


FIG. 3. Rough isodose curves drawn in location of tilted final viewing screen, with a 50-mil condenser aperture and sufficient intensity for taking motion pictures. Values would be much higher from intermediate screen.

the films very instructive. The same is true in lectures on electron microscopy where an audience is given an improved impression if a movie sequence is shown as illustration, even when the motions recorded are themselves not particularly informative for new things. The ability to reverse or retard a film is very useful, especially when observing the extremely fast reactions which occur during the earliest stages of a bombardment series.

Motion Picture Studies of Electron Bombardment of Colloidal Crystals

JOHN H. L. WATSON AND LUTHER E. PREUSS
Edsel B. Ford Institute for Medical Research, Detroit 2, Michigan

(Received April 12, 1950)

Phenomena not seen or recorded by orthodox methods are observed on the large number of micrographs which are taken during a motion picture experiment, and reaction times as short as 0.03 sec. can be calculated. The technique is useful for investigating solid-liquid-solid changes both by direct observation and by recording Bragg reflections. In most cases a preliminary effect of electrons upon crystals is to drive off water. In the case of sodium chloride crystals a residual envelope is left. Inside this, small particles, both solid and liquid, are observed to move rapidly due, it is believed, to a combination of Brownian movement and convection. These gradually evaporate and diffuse through the wall. The possibility is discussed that the residual envelope is formed by contamination. Examples are also taken from bombarded, colloidal crystals of tungsten oxide.

GERMAN workers realized at an early date the advantages to be gained in electron microscopy by the use of a battery-biased source, and described some of the effects upon their specimens of the intense beam from such a gun. The major emphasis was placed upon the subject in this country after Hillier and Baker[1] described the self-biasing cathode. Several reports[2–4] have been published since which discuss from stills the effects of electrons upon matter as observed in electron microscopes employing self-biasing sources, and observations of these reactions are often used now to clarify electron microscopic sample interpretation.

Several colloidal crystal systems have been studied by application of the motion picture technique[5] for their reactions under electron radiation. The first of these involves crystals of sodium chloride.[6] In Fig. 1 several stills are shown which are typical of a movie sequence. It will be seen that the contrast of the "moving micrographs" is good and that the resolution and magnifications are useful, if not completely comparable with the best to be expected from the more orthodox methods of electron microscopy. Figure 1A shows an early stage of the bombardment series as the water is being driven off and sections of the crystal are becoming transparent to electrons. Before this the whole crystal was quite opaque. Figures 1 (B–F) are all progressive, intermediate appearances of the same crystal as it changes toward the residual envelope[6] condition which is its final state, Fig. 1G.

The fine particles, which are seen to form within the envelope, move very rapidly in a random fashion under bombardment. This motion slows or ceases whenever the intensity is diminished. From higher resolution micrographs taken in the orthodox manner, Fig. 2, these particles are often seen to be rounded, but in many cases are small cubes. The rounded particles are apparently molten and are frequently noticed to coalesce with others which they meet. The cubes are presumed to be cooler, solid crystals. All the particles continue to decrease in size and contrast until eventually they evaporate completely, diffuse through the envelope as vapor and disappear.

The motion of the particles is probably a combination of convection and Brownian movement of the particles within the confined space of the envelope under the influence of the warm, vaporous material which fills it. That it is filled with such a material under pressure is shown by the facts (1) that the sides of the envelope will often bulge outward and even break if heat is applied too rapidly and (2) that such a break results in the spewing out of vaporous substance from the envelope which deposits as solid particles on the cool, surrounding substrate. These effects are demonstrated very graphically in the moving pictures. It has been thought that Brownian motion would not be studied to advantage in an electron microscope due to the extremely large translations which would be involved at the high magnifications on the very limited viewing screen. Here, however, the motion is constrained within narrow limits by the residual envelope. Figure 3 from another movie sequence shows a ruptured envelope.

The residual envelope has been conjectured[6] to originate from contamination[7] but if so, the contamination must be much more quickly and effectively deposited upon some materials than upon others, for in these studies only sodium chloride has been found to leave such envelopes. This observation agrees with the results of Burton *et al.* who classed sodium chloride along with sodium chlorate, potassium bromide, and iodide as substances which behaved in this way. If contamination explains the occurrence of residual envelopes, this specificity for certain substances is in disagreement with the qualitative observations by Hillier[8] that the rate of deposition was the same for all substances. On the other hand, if Hillier's observation is correct, other explanations must be looked for to explain the presence of

[1] James Hillier and R. F. Baker, J. App. Phys. **17**, 12 (1946).
[2] John H. L. Watson, J. App. Phys. **19**, 713 (1948).
[3] Heller, Wojtowicz, and Watson, J. Chem. Phys. **16**, 998 (1948).
[4] F. A. Hamm, and Earl Van Norman, J. App. Phys. **19**, 1097 (1948).
[5] Luther E. Preuss and John H. L. Watson, J. App. Phys. **21**, 902 (1950), preceding paper.
[6] Burton, Sennett, and Ellis, Nature **160**, 565 (1947).
[7] John H. L. Watson, J. App. Phys. **18**, 153 (1947).
[8] James Hillier, J. App. Phys. **19**, 226 (1948).

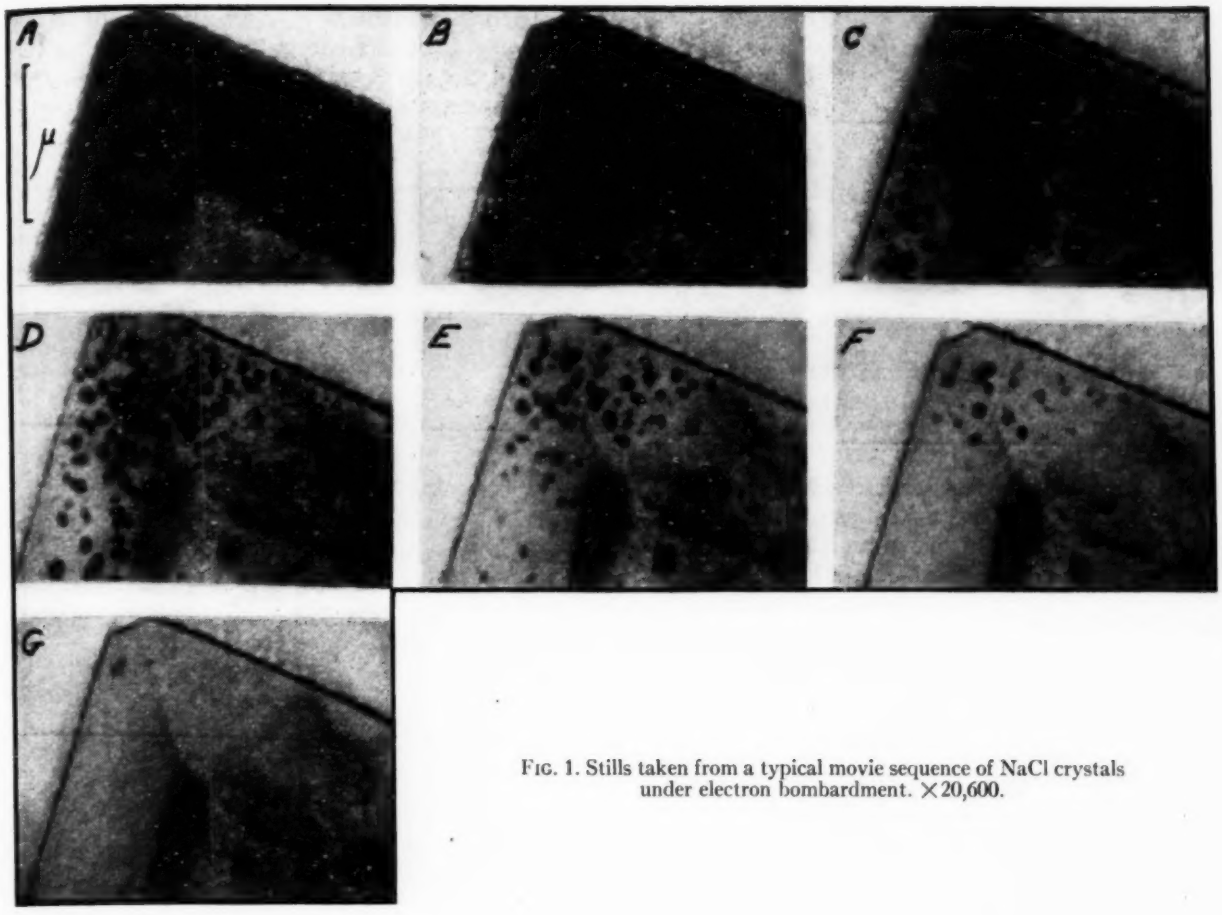


FIG. 1. Stills taken from a typical movie sequence of NaCl crystals under electron bombardment. ×20,600.

residual envelopes. The envelope is always present with sodium chloride regardless of the speed of evaporation, which often occurs instantaneously. It may be significant that those salts which are water soluble tend to leave residual envelopes while the insoluble ones do not. It is interesting to hypothesize that the residual layer is left as a result of trace impurities which concentrate at the original surface during the evaporation of the interior. Salt crystals deposited from saline, organic suspensions yield particularly striking envelopes, Fig. 4, a fact which may indicate that organic impurities are polymerized or condensed at the crystal surface in a radiation-chemical process under the action of the electron beam.

The three-dimensional nature of the residue as a replica of the original crystal has been demonstrated by shadow-casting them after evaporation and by tilting them on broken Formvar films and recording their appearance in movie film. As a replica they give information concerning the surface of the original crystals.

The large number of micrographs taken rapidly by the moving picture technique has resulted in several examples being recorded of a transient appearance of an inherent mosaic structure in sodium chloride crystals, in support of the discussion by Ewald and Renninger[9] that a mosaic exists. The phenomenon was observed to form and disappear quickly, and its presence was detected in only one or two stills of any sequence in which it ap-

FIG. 2. Still taken by orthodox, electron microscope procedure of NaCl crystal under electron bombardment. Reversed print. ×12,000.

[9] P. P. Ewald and M. Renninger, International Conference on Physics Papers and Discussions (Physical Society, 1935), Vol. 11, pp. 57–61.

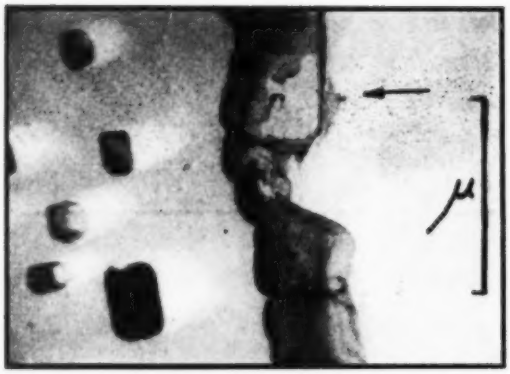


FIG. 3. Still taken from a typical movie sequence showing a ruptured envelope at arrow. ×26,000.

peared. Its visibility was therefore a short-lived affair which would not be micrographed as successfully without the use of the moving picture methods. One still is shown in Fig. 5 as an illustration. It is an interior rather

FIG. 4. Electron micrograph of evaporated salt crystals in a bacterial preparation. Reversed print ×3600.

than a surface effect, since it disappears under continued bombardment.

The bombardment effects of electrons upon tungsten oxides (WO_3) have been reported earlier,[1,2] but are more

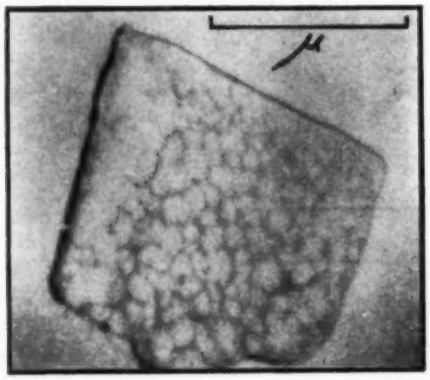


FIG. 5. Still taken from a typical movie sequence, showing the mosaic structure in an evaporating NaCl crystal. ×26,000.

completely explainable from observations made on film. By slowing down the effects considerably it has been confirmed that the mosaic pattern of rods in fresh, early samples is inherent. The initial effect of bombardment is to render these visible but not to create them. Apparently, the heat from the incident beam drives off water of crystallization and some free passage of electrons results. The initial stages of bombardment of these and of most other crystals passes extremely rapidly, depending upon the intensity, and these stages are the most productive in moving picture studies. The effects of melting and recrystallization into other forms are secondary and enter later.

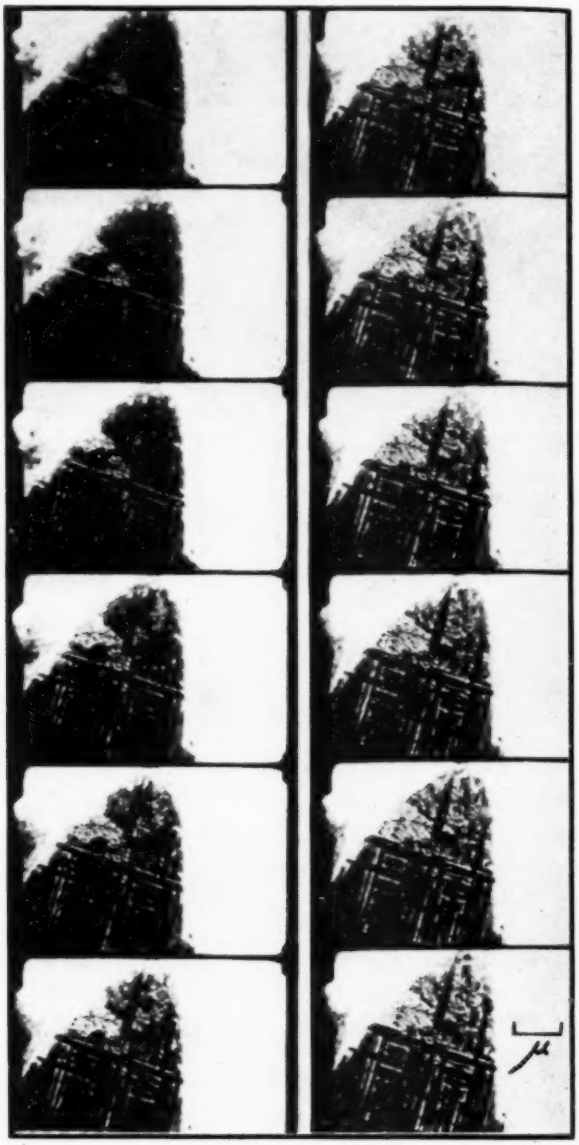


FIG. 6. A typical sequence of frames taken of the effects of electron bombardment upon WO_3. The sequence was taken at 16 frames per second and covered a total time interval of $\frac{3}{4}$ sec. ×7200.

Figure 6 shows a typical sequence of frames taken at an original magnification of ×7200 in the microscope and reproduced here at the same magnification. Since these were taken at 16 frames per second, the observed reactions occurred within a very short interval. The actual exposure time per frame for this example was 1/30 sec., and the whole sequence represents a time interval of ¾ sec. In the first one or two frames little if any reaction has occurred although some of the inherent structure is seen to be giving rise to Bragg reflections which appear as a straight, intense line over the crystal surface. Later the inherent mosaic has become visible over the whole body of the somatoid. In the final frames secondary structures due to melting have been introduced at the tip of the crystal.

There is no suggestion of a residual envelope when tungsten oxide is bombarded. Melting is recognizable from the round nature of the melted particles, and recrystallization is interpreted from production of moving Bragg reflections and from a return to the straight-edged morphology. Areas presumed to be melted often give rise to electron reflections. Melted particles are also seen to "bubble" in the moving micrographs as if boiling under the beam.

Moving pictures of carbon black contaminating under the beam demonstrated the particle growth very graphically, but added no new knowledge of the phenomenon.

The Emission of Radiation from Nitric Oxide: Approximate Calculations*

L. E. Benitez** and S. S. Penner
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California
(Received March 27, 1950)

Emissivity calculations have been carried out for nitric oxide at temperatures from 300°K to 3000°K. The results may be used, as a first approximation, at elevated total pressures.

NITRIC oxide[1] is one of the diatomic molecules which may contribute to radiant heat transfer in high pressure combustion chambers. In particular, nitric oxide at elevated temperatures is encountered in many solid-fuel and liquid-fuel rocket motors. It is, therefore, of obvious practical importance to determine the emissivity of nitric oxide as a function of temperature. According to the procedure for making approximate radiant heat-transfer calculations outlined previously,[2] it is necessary to calculate first an effective band width lations are then used to construct a plot of limiting emissivities for the fundamental and the first overtone of NO as a function of temperature. Finally, emissivity calculations are made by using the limiting emissivities in conjunction with the values of the integrated absorption for the fundamental and the first overtone, estimated from available experimental absorption measurements of NO.

The results of effective band-width calculations, using the procedure described previously[2] and the spectroscopic constants given by Mayer and Mayer,[3] are summarized in Tables I and II for the fundamental and the first overtone of NO, respectively, for two different values of $j_{max.}$ at each temperature. It should be noted that the presence of a Q branch in the infra-red spectrum of NO does not affect the effective band-width calculations made according to our approximate procedure. For values of the limiting rotational quantum number $j_{max.}$ exceeding the value corresponding to the position

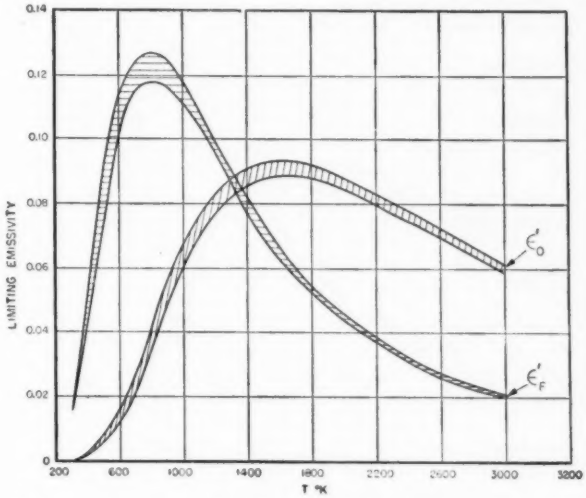


FIG. 1. Limiting emissivity for the fundamental and first overtone of NO, as a function of temperature, for two sets of values of $j_{max.}$.

* This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. W-04-200-ord-455, sponsored by the U. S. Army Ordnance Department.

** Lieutenant, U. S. Navy. Present address: Massachusetts Institute of Technology, Cambridge, Massachusetts.

[1] Further details concerning radiant heat-transfer calculations on nitric oxide may be found in a thesis submitted by Lt. L. E. Benitez in partial fulfillment of requirements for the degree of Aeronautical Engineer, California Institute of Technology, June 1949.

[2] S. S. Penner, J. App. Phys. **21**, 685 (1950).

[3] J. E. Mayer and M. Goeppert-Mayer, *Statistical Mechanics* (John Wiley and Sons, Inc., New York, 1940), p. 469. The numerical results in the present report are practically unchanged if the spectroscopic constants given by R. H. Gillette and E. H. Eyster, Phys. Rev. **56**, 1113 (1939), are used.

TABLE I. Effective band width of the fundamental of NO as a function of temperature.

T °K	$j_{max.}$	$\nu_{max.}$ cm^{-1}	$\nu_{min.}$ cm^{-1}	$\Delta\nu_{n\to n+1}$ cm^{-1}
300	30	1962	1761	201
	33	1968	1748	220
1000	54	2003	1646	357
	58	2007	1625	382
1500	65	2013	1587	426
	69	2015	1566	449
2000	77	2018	1519	499
	81	2019	1496	523
2500	88	2019	1454	565
	91	2019	1436	583
3000	100	2019	1379	640
	106	2019	1341	678

TABLE II. Effective band width of the first overtone of NO as a function of temperature.

T °K	$j_{max.}$	$\nu_{max.}$ cm^{-1}	$\nu_{min.}$ cm^{-1}	$\Delta\nu_{n\to n+2}$ cm^{-1}
300	30	3793	3594	199
	33	3796	3577	219
1000	54	3801	3442	359
	58	3801	3412	389
1500	65	3801	3360	441
	69	3801	3328	473
2000	77	3801	3260	541
	81	3801	3226	575
2500	88	3801	3162	639
	91	3801	3134	667
3000	100	3801	3046	755
	106	3801	2985	816

of a band head, the limiting frequency has been set equal to the frequency at the band head.[2]

From the data listed in Tables I and II the limiting emissivities have been calculated. The limiting emissivities ϵ_F' and ϵ_0' for the fundamental and the first overtone, respectively, are defined by the relations

$$\epsilon_F' = \int_{\Delta\nu_{n\to n+1}} \rho(\nu)d\nu \Big/ \int_0^\infty \rho(\nu)d\nu,$$

$$\epsilon_0' = \int_{\Delta\nu_{n\to n+2}} \rho(\nu)d\nu \Big/ \int_0^\infty \rho(\nu)d\nu,$$

where $\rho(\nu)$ is the energy density of radiation emitted by a blackbody at a given temperature. The limiting emissivities for the two sets of $j_{max.}$ values are plotted as a function of temperature in Fig. 1.

In order to utilize the plot of limiting emissivities shown in Fig. 1 for radiant heat-transfer calculations, it is necessary to estimate the integrated absorption for the vibration-rotation bands.[2] Reliable numerical values do not appear to be available for NO although qualitative results have been reported by Gillette and Eyster[3] and by Rideal *et al.*[4] Experimental measurements of apparent absorption coefficients of NO and CO at small total pressures with an instrument of low spectral resolution have led to similar experimental results for the integrated absorption for these two gases.[1] On the basis of these studies it is suggested that the emissivity ϵ' be determined by use of the relations

$$\left.\begin{aligned} \epsilon' &= [1-\exp(-\bar{k}_F pl)]\epsilon_F' \\ &\quad +[1-\exp(-k_0 pl)]\epsilon_0', \\ k_F(T) &= \frac{201}{\Delta\nu_{n\to n+1}}\times\frac{300}{T}\times 116 \\ &= \frac{7.0\times 10^6}{T\Delta\nu_{n\to n+1}}, \text{ (cm-m-atmos.)}^{-1}, \\ k_0(T) &= \frac{200}{\Delta\nu_{n\to n+2}}\times\frac{300}{T}\times 0.88 \\ &= \frac{5.3\times 10^4}{T\Delta\nu_{n\to n+2}}, \text{ (cm-m-atmos.)}^{-1}, \end{aligned}\right\} \quad (1)$$

where the following numerical values have been used at 300°K:

$$\Delta\nu_{n\to n+1} = 201 \text{ cm}^{-1},$$

$$\Delta\nu_{n\to n+2} = 200 \text{ cm}^{-1},$$

$$\bar{k}_F = 116 \text{ (m-atmos.)}^{-1},$$

$$k_0 = 0.88 \text{ (m-atmos.)}^{-1}.$$

The numerical values of $\bar{k}_F$ and $\bar{k}_0$ should be regarded as rough approximations subject to revision on completion of current measurements of integrated absorption.

[4] Snow, Rawlins, and Rideal, Proc. Roy. Soc. (London) **124A**, 453 (1929).

Generalized Microscopy and the Two-Wave-Length Microscope

M. J. Buerger
Crystallographic Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received April 27, 1950)

In the ordinary simple microscope, the same light is permitted to flow through the entire system. A more general microscope can be devised in which light of one wave-length is permitted to flow as far as the diffraction image of the object, then light of a second wave-length is substituted at this plane, and continues to flow through the rest of the optical system. It is shown that for such a generalized microscope, the magnification depends not only on the ordinary image-to-object distance ratio, but also on the ratio of wave-lengths used. Utilizing x-rays and visible light, this second factor is about 10^4 diameters, and the whole magnification is of the order of 3×10^5. This is enough to permit one to see an atom. Apparatus has been built which realizes this theory. A most important characteristic of any such system is that a loss of phase occurs at the first diffraction image, where the wave-length substitution takes place. This phase must be supplied by the apparatus. It is accomplished with the aid of phase shifters prepared from a large uniform mica cleavage fragment. The apparatus and the results achieved with it are described.

THE generalized theory of microscopy discussed here was developed by the writer in 1939, in connection with the theory of optically reciprocal gratings,[1] and it was presented at the January 10, 1941 meeting of the New York Academy of Sciences, at the October 20, 1942 meeting of the Rochester Section of the American Optical Society, and on a number of subsequent occasions. The subject matter is, therefore, not entirely new. But what makes it timely and interesting is that a technical development which was in a successful experimental stage in 1941 has now been perfected, and therefore new and very interesting results have recently been achieved with this instrument. The gap in development was due to the intervention of the war, and consequent diversion of interests.

The two-wave-length microscope is a device for magnifying an image in two stages, using a different wave-length in each stage. The magnification then depends not only on the usual ratio of image-to-object distances, but also on the ratio of the wave-lengths used. To understand this action, consider first an ordinary optical system employing the same wave-length of light throughout, shown in Fig. 1. This consists of two lenses L_1 and L_2, of focal lengths D_1 and D_2, respectively, and separated by a distance D_1+D_2. An illuminated object is placed at the left focus of the first (left) lens, and its image is formed at the right focus of the second (right) lens. The dotted lines indicate the geometrical optics of image formation.

Now, this simple process can also be regarded as two stages of diffraction in series. The first lens, with the object at its focus, collects parallel rays emanating from the object and focuses these at its right focal plane. This is the diffraction image of the object. Then those rays emanating from this diffraction image which are parallel are focused by the second lens to become the final image. Thus the final image is also the *diffraction of the diffraction* of the object. The focusing of the diffraction in these two stages by the two lenses is indicated in Fig. 1 by the solid lines.

If the object is periodic, then its diffraction image is in the form of a collection of discrete spectra. Let the periodicity of the object be defined by the lattice of the object. Then,[1] as will be shown, the diffraction image is the weighted reciprocal lattice[2] of the object. The diffraction image of this reciprocal lattice is the final image. To generalize this statement for non-periodic objects, one can restate it in the language of Fourier transforms. The diffraction of the object is its Fourier transform. The diffraction of the diffraction image is therefore the *transform of the transform* of the object. It is well known that this is the enantiomorphous equivalent of the object.

Consider the magnification of the image in terms of this double diffraction process. It simplifies the treatment to carry through the discussion of magnification for a periodic object, thus avoiding the explicit use of Fourier transforms. Let the translations t_1 and t_2 define the periodicity of the pattern of the object in its own plane. Then the wavelets scattered by the pattern units of the object are alternatively in phase and out of phase in a plane at right angles to a set of parallel rows, Fig. 2. If the spacing of such rows is d, then the condition that the wavelets are in phase (Fig. 3) is

$$\sin\delta = n\lambda_1/d. \qquad (1)$$

These re-enforcing waves are focused on the plane of the diffraction image at a point whose distance, t_n^*, from

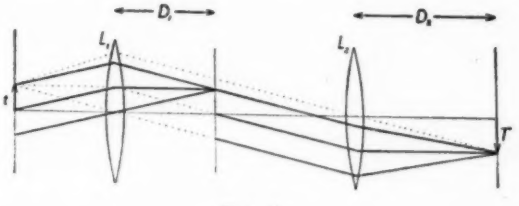


FIG. 1.

[1] M. J. Buerger, "Optically reciprocal gratings and their application to the synthesis of Fourier series," Proc. Nat. Acad. Sci. **27**, 117–124 (1941).

[2] M. J. Buerger, *X-Ray Crystallography* (John Wiley and Sons, Inc., New York, 1942), pp. 107–127.

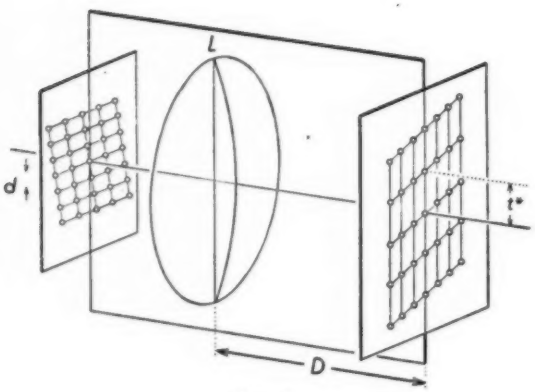


FIG. 2.

the optic axis of the system (Fig. 3) is given by

$$\tan\delta = t_n^*/D_1. \tag{2}$$

Combining (1) and (2),

$$t_n^* = D_1 \tan \sin^{-1}(n\lambda_1/d). \tag{3}$$

Provided that δ is small (i.e., that discussion is limited to spectra comparatively near to the optic axis), a close approximation is

$$t_n^* = n(D_1\lambda_1)/d. \tag{4}$$

Thus, the collection of the n spectra constitutes a row of points whose interval is

$$t^* = (D_1\lambda_1)/d. \tag{5}$$

Such a row of spectra occurs normal to *any* set of rows in the object Fig. 2, i.e., they occur parallel to the d_{hk} of any set of rows of crystallographic index (hk) and the law of the interval t_{hk}^* in the spectrum is

$$t_{hk}^* = K/d_{hk}, \tag{6}$$

where K is an experimental constant consisting of the product $(D_1\lambda_1)$.

A collection of points displaying this kind of reciprocity with respect to the spacings of all the rows of a periodic object is known to be a lattice,[2] and in fact it is the reciprocal lattice of the object. Thus the diffraction of a periodic pattern is a set of discrete spectra at points of the reciprocal lattice of the object.

But this reciprocity is deeper than merely the reciprocity in the dimensions of t^* and d; the reciprocity is truly reciprocal. That is, the star can be transferred from the d to the t, and therefore t and d^* are reciprocal, i.e., the translations of the lattice of the object and the spacings of the diffraction image are also reciprocal. That this is so can be quickly appreciated by imagining the direction of the light in Fig. 1 to be reversed. The diffraction from the reciprocal lattice is then the object lattice. Thus it is also true that

$$t_{uv} = (D_1\lambda_1)/d_{uv}^*. \tag{7}$$

(In the above discussion, the starred letters always refer to quantities in the diffraction image, i.e., to the reciprocal lattice.)

Furthermore, it is evident that, since the arrangement of spectra in the diffraction image is on points of a lattice, the forward diffraction of the diffraction image is also periodic. The same law as derived above holds, and the length of a period of the final image corresponds to that derived for a period of the diffraction image, but specifically

$$T = (D_2\lambda_2)/d^*. \tag{8}$$

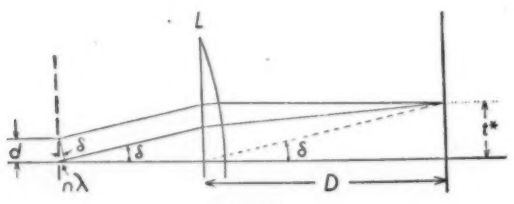


FIG. 3.

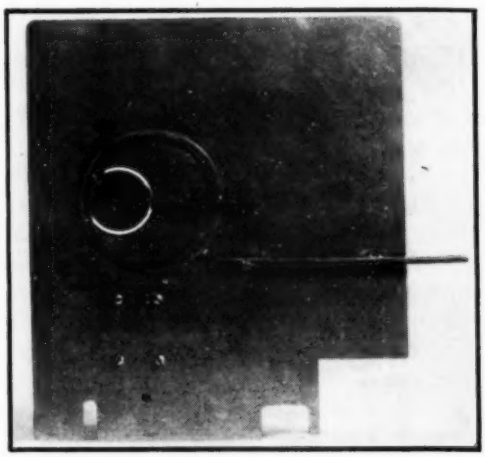


FIG. 4. Device for removing the Lorentz and polarization factors during the recording of de Jong-Bouman photographs. The large metal sheet with the annular opening acts as the layer-line screen. A sector rotates at high speed in front of the opening, propelled by an air jet. The interposing of the metal of the sector in the path of the reflections reduces the time during which each one records, by a factor which is proportional to the product of the Lorentz and polarization factors for the zero level.

The over-all magnification of the optical system is evidently

$$M = \frac{\text{length of a period in the image}}{\text{length of a period in the object}} = \frac{T}{t}. \tag{9}$$

Substituting in (9) from (8) and (7), the magnification turns out to be

$$M = (D_2/D_1) \times (\lambda_2/\lambda_1). \tag{10}$$

Now, in the usual optical system, the same light is used for both stages of diffraction, and (10) reduces to $M = D_2/D_1$, which is the usual ratio of image distance to object distance. But tremendous magnifications could be achieved if two very different wave-lengths

were used for the two stages of diffraction. Specifically, if x-rays were used for the first stage, and visible light for the second stage, a factor of 10^4 could be realized apart from the further magnification due to choice of focal lengths of the lenses. Some specific values of the factors in the right of (10) which are used in the instrument described later are approximately

$D_2 = 2\times10^2$ cm
$D_1 = 5$ cm
$\lambda_2 = 5\times10^{-5}$ cm (green light)
$\lambda_1 = 7\times10^{-9}$ cm (molybdenum $K\alpha$-radiation).

With these values, (10) becomes

$$M = \frac{2\times10^2}{5} \times \frac{5\times10^{-5}}{7\times10^{-9}} = 3\times10^5 \text{ times.}$$

With this magnification, an atom 1A in diameter would appear as a circle 0.03 mm in diameter, which is large enough to see with a low power microscope. Such a scheme of magnification would have the advantage that *the resolving power of the system depends on the first wave-length, and consequently this enormous magnification is not empty.*

This system has two inherent disadvantages. In the first place, unless the first radiation of wave-length λ_1, arriving at the diffraction image, can be caused by some device to release the second radiation of wave-length λ_2 in appropriate phase, then the phase of each spectrum

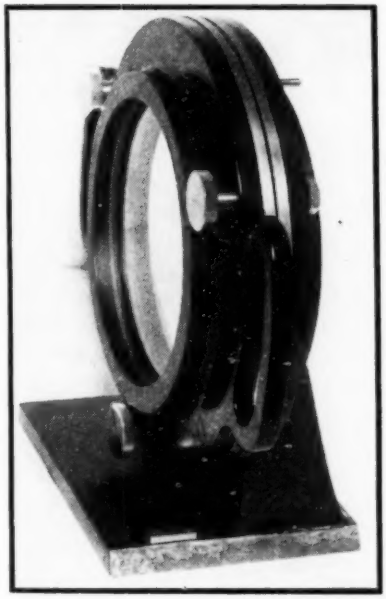


FIG. 5. Device for utilizing a print on film as a diffraction grating. The print is sandwiched between two thick glass flats, and the whole thickness is then rendered optically homogeneous by filling all space between film and flats with a liquid whose refractive index matches that of the film. The liquid is run into the system through the tube seen projecting at the upper right. The (forward) movable flat can be grasped by the transverse handles and removed after releasing the three screws. The clear aperture of the system is about 6 in.

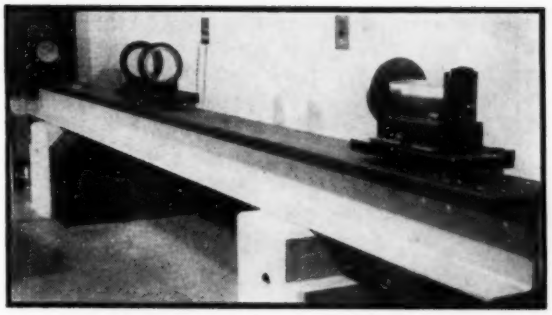


FIG. 6.

of the diffraction image is lost. At present there appears to be no obvious device for accomplishing this phased release of the second radiation. In the absence of such a device, the two-wave-length microscope requires that the phases be supplied. The mechanics of supplying the phases are described later.

Another inherent disadvantage of this system is that if a tri-periodic pattern, such as a crystal, is used as an object, the conditions for diffraction by a three-dimensional grating obtain. One aspect of such diffraction is that, in general, the spectra can only be obtained one at a time as the object is shifted in orientation to satisfy the Bragg condition for each plane one at a time. Thus, the diffraction from the diffraction image could not be made to interfere and build up a final image. Furthermore, if x-rays are to be used in the first stage, they cannot be focused by lenses. But fortunately, both of these difficulties can be avoided because there exist instruments for recording x-ray diffraction in a pattern which is arranged precisely as if the pattern *had* been focused by lenses and as if the several diffraction spectra *had* been produced at once. This is accomplished by the de Jong-Bouman apparatus[3] or by the precession apparatus.[4] The law of magnification for these instruments is precisely (5), where D_1 is the crystal-to-film distance, and λ_1 is the x-ray wavelength.

Thus, in effect, the system *can* be used to form a magnified image of the patterns of crystals provided the appropriate phases of the diffraction by the diffraction image can be supplied. To use the system, it is only necessary to substitute at the plane of the first diffraction image, the positive photograph of the x-ray diffraction pattern (or its equivalent) arranged in reciprocal lattice form by one of the devices mentioned in the last paragraph, and then illuminate this pattern with visible light. In this way, in effect, visible light is substituted for x-rays in the right part of Fig. 1. The philosophy of generalized microscopy is then this: Record (for example, by photography) the diffraction image of the first stage of image formation using radia-

[3] See reference 2, pp. 331–346.
[4] M. J. Buerger, "The photography of the reciprocal lattice," Am. Soc. X-Ray and Electron Diffraction, Monograph No. 1, pp. 1–37 (1944).

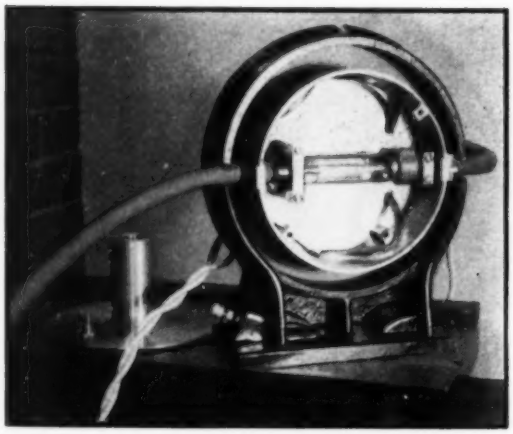


FIG. 7(a). Water-cooled mercury arc source.

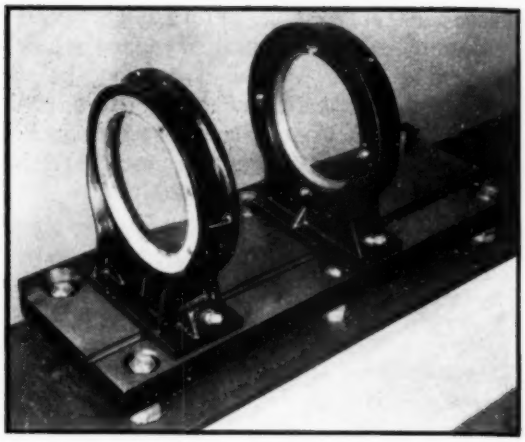


FIG. 7(c). Lens assembly.

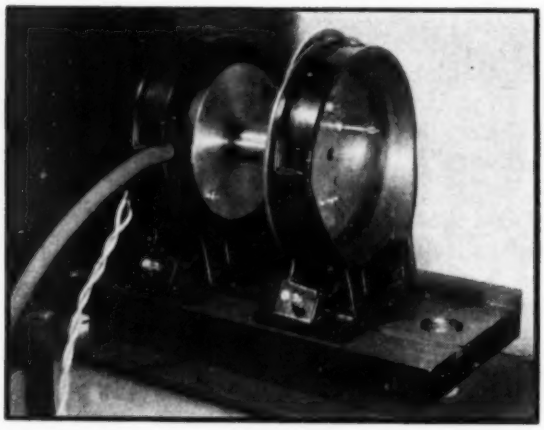


FIG. 7(b). Mercury arc source focused on a pinhole of the pinhole assembly.

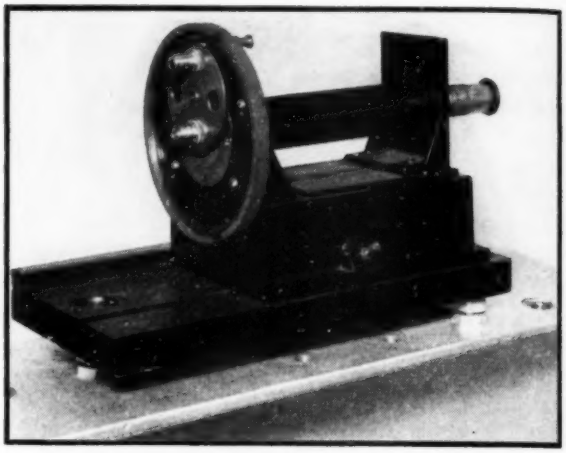


FIG. 7(d). Microscope assembly.

tion of short wave-length; then, placing the positive photograph (or equivalent) of this diffraction image at the correct position of the diffraction image, illuminate it with radiation of longer wave-length. After providing the right phases, the image now obtained is the same as if the original radiation had been used throughout, except that the magnification is scaled up by a factor λ_2/λ_1.

The use of positive photographs of x-ray diffraction arranged in reciprocal lattice form has been practiced by the writer. It has several disadvantages. In the first place, the x-ray diffraction cameras distort the intensities of the spectra by multiplying each by the Lorentz and polarization factors.[5] Fortunately this distortion can be automatically removed during recording by a rapidly rotating sector,[5] shown in Fig. 4. Second, the photographic film surfaces are too irregular for diffraction work. This can be offset by immersing the film in a liquid of its own refractive index, and sandwiched between two optical flats utilizing apparatus such as shown in Fig. 5. This has also been successfully practiced.

But, in spite of its inelegance, a more satisfactory way, from a mechanical point of view, is to determine the intensities of the x-ray diffraction spectra by the Dawton technique[6] (correcting for Lorentz and polarization factors if this has not already been done automatically during recording) and make a mechanical equivalent of the positive diffraction pattern consisting of a metal plate with a hole at each spectral position whose area is proportional to the magnitude of the amplitude $|F_{hkl}|$ of the spectrum. A large number of unit cells of the structure can be accurately imaged in the focused diffraction of such a grating if the largest hole does not exceed about 50 wire gauge (0.070 in. diameter). This equivalent of the first diffraction image has mechanical qualities which lend themselves well to the phase-shifting arrangements to be described.

A view of the assembled optical units of the apparatus which represents the second diffraction stage of Fig. 1

[5] M. J. Buerger, "The correction of x-ray diffraction intensities for Lorentz and polarization factors," Proc. Nat. Acad. Sci. **26**, 637–642 (1940).

[6] M. J. Buerger, "The photography of interatomic distance vectors and of crystal patterns," Proc. Nat. Acad. Sci. **25**, 383–388 (1939).

is shown in Fig. 6. This comprises, from left to right, a source of light, a pinhole, two lenses, and a microscope. The light source illuminates the pinhole, which acts as a point source of light. The diverging radiation from this source is collimated by the first lens and strikes the reciprocal-lattice diffraction grating as a plane wave. The diffraction from the grating is focused by the second lens so that the final image occurs near the right side of the assembly. The device at the extreme right is a low power microscope which magnifies the image so that its details are easily visible to the eye.

The units of the optical system are shown in Fig. 7. They are mounted on individual beds. Each unit and bed is fitted with a millimeter scale and vernier attachment so that the unit, if removed for any purpose, can be accurately returned to its correct location.

The optical path of the system is normally shielded from the disturbing influences of stray light by a tubular cover, seen in Fig. 8. The optical system is also fitted with an auxiliary mirror which can be raised into the cone of rays converging from the second lens. This throws the diffraction image back to a point near the grating, where it can be viewed by the operator from this region, Fig. 9. This arrangement permits the operator to make any desired adjustments on attachments to the grating, such as the phase shifters to be described beyond, while watching the results.

The source of light is a water-cooled mercury arc, specifically the General Electric *A-H*6. The mercury green line (5461A) is isolated from the arc spectrum. Shorter wave-lengths are eliminated by a Wratten filter No. 12, whose cut-off is nearly complete below about 4900A. This is placed permanently behind the lens of the microscope objective. Longer wave-lengths, particularly the orange component near 5800A, are eliminated by taking the photographs on orthochromatic film.

The mercury arc illuminates a very small pinhole. The limitations on the size of this pinhole can be appreciated if one notes that, with the magnification mentioned earlier, an atom one angstrom in diameter would appear as a disk about 0.03 mm in diameter. This is about 0.001 in., and the upper limit of the diameter of the pinhole must not substantially exceed this, and preferably it must be considerably less in order to resolve detail in the individual atoms of the image. In the apparatus shown in Fig. 6 there is a turret device, Fig. 7(b), which permits one to swing pinholes of four different sizes into position. The two largest pinholes are 0.0015 and 0.0007 in. in diameter, respectively. These were prepared by drilling holes mechanically in thin pieces of brass sheet. Their principal use is to examine images for atomic position, where atomic detail is not required. The turret device also bears two very much smaller pinholes of graded sizes which were found to occur fortuitously in thin aluminum foil. These are necessary when detailed photographs, such as shown in Fig. 13, are to be taken.

The two lenses are plano-convex and are substantially duplicates, with diameters of 6 in. and focal lengths of 6 ft. They are very perfectly corrected for spherical abberation but, of course, require no correction for chromatic abberation since they are used with monochromatic light.

The microscope at the right end of Fig. 6 has two standard parfocal microscope objectives of 16 and 32 mm focal lengths, respectively, mounted in a special turret so that they can be quickly interchanged, Fig. 7(d). The microscope image can be either examined visually through an eyepiece, or a small film holder can be substituted for the eyepiece, Fig. 6, and the image photographed. As mentioned above, orthochromatic film is used for this purpose in order to eliminate the orange mercury line.

It has been pointed out[1] that any phase shift whatever can be imposed on the rays scattered by a point of the reciprocal grating by interposing in the path of the rays going through the point a tilted flake of mica which has been ascertained to have a truly uniform thickness. This was experimentally proven by the writer to be feasible as early as 1941. For this purpose, a thin cleavage flake having as large an area as possible is separated from an excellent crystal of mica. The thickness uniformity is easily judged by a very simple interferometer method: The flake is merely held close to a strong broad source of monochromatic light, such as a General Electric "Labarc." Interference fringes then appear to contour the mica flake, and these are without discontinuities in fields where the cleavage flake has a uniform thickness. It is a simple matter by

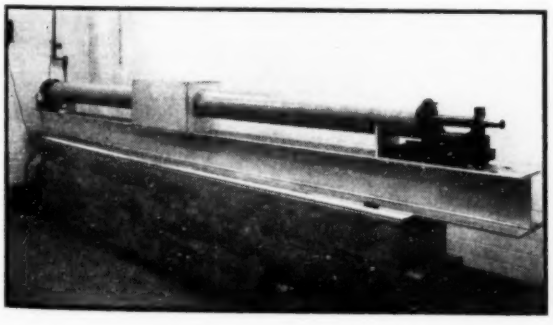


FIG. 8.

FIG. 9.

FIG. 10.

such observations to map out the regions of the flake which are uniform in thickness. A large uniform area is then selected for the purpose of cutting out the phase shifter units described below.

In order to calibrate the mica sheet so prepared, a small sample of the sheet is cut out and mounted on a rotation device which is placed so that the mica is directly in front of one of two equal circular holes of a two-hole "grating," shown in Fig. 10. The diffraction pattern of the pair of holes can be examined by placing the "grating" in the optical apparatus described above. Without the mica in the system, the diffraction pattern consists of a series of short linear maxima whose long directions are normal to the separation direction of the two holes. When the mica sample is placed in front of one of the holes, these maxima are shifted, in general, in accordance with the phase shift induced by the additional path length through the mica. This path length, and its attendant phase shift, is a function of the angle

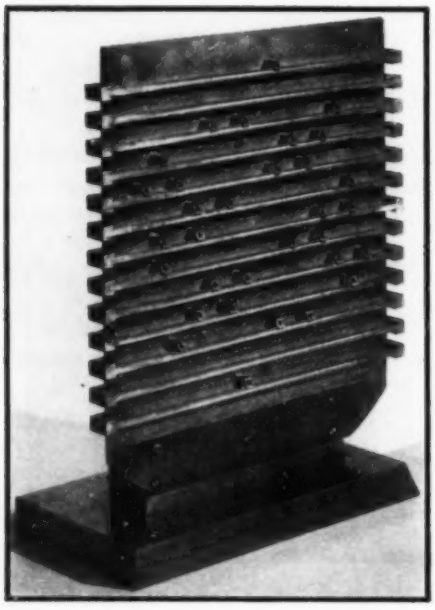


FIG. 11.

which the normal to the mica sheet makes with the rays passing through the hole. By measuring the drift of the fringes as a function of the mica angle, the phase shift produced by a particular mica flake can be determined as a function of this angle. For example, the setting of the mica which causes a shift of maxima so that one maximum comes to be located where a minimum occurred with no mica in place, corresponds to a phase shift of π. Proceeding in this way, the angle which the mica should make with the optical axis of the system to cause a phase shift of any required amount can be determined. The theory and practice of constructing phase shifters will be discussed in more detail in another paper.

A practical way of placing the mica sheets at the correct angles in front of the various holes of the grating is illustrated in Fig. 11. This particular example shows phase shifters in place in front of the grating used for producing the image of marcasite, FeS_2, described later and shown in Fig. 13. The individual mica pieces are mounted on small brass blocks in such a way that the

FIG. 12. Left to right: (a) Assembly for producing phase shift of π for centrosymmetrical images. (b) Assembly for producing any phase shift required for non-centrosymmetrical images. (c) and (d) The assembly shown in (b) taken apart. The piece in the right is the mica-bearing member whose orientation can be varied within the metal block seen in (c).

mica makes the required angle with the rays through the grating holes. The blocks are placed on shelves along each row of one set of rows in the reciprocal lattice.

It is convenient to have two kinds of phase-shifter blocks. The rather simplified block shown in place in Fig. 11 is desirable when the pattern of the crystal structure projection to be imaged is centrosymmetrical. In such cases the phases can have only the values 0 and π, so only two discrete values of phase must be impressed on the scattering from the holes of the grating. The phases of the waves scattered by those holes whose scattering phase is to be set at zero need not be regulated, for the phase of the plane wave impinging on the grating may be taken as zero. But the phases of the waves scattered by holes whose scattering phase is to be set at π must be controlled by placing in front of each such hole a block bearing a piece of mica set at the correct angle to produce the phase shift of π. The design of this simple phase shifter is shown in Fig. 12. To image an average crystal structure, about 100 phase shifters are ordinarily needed. This requires that the original piece of mica should be of sufficiently large area to permit cutting 100 duplicate pieces from it. Each such fragment is cut so that the rotation of the

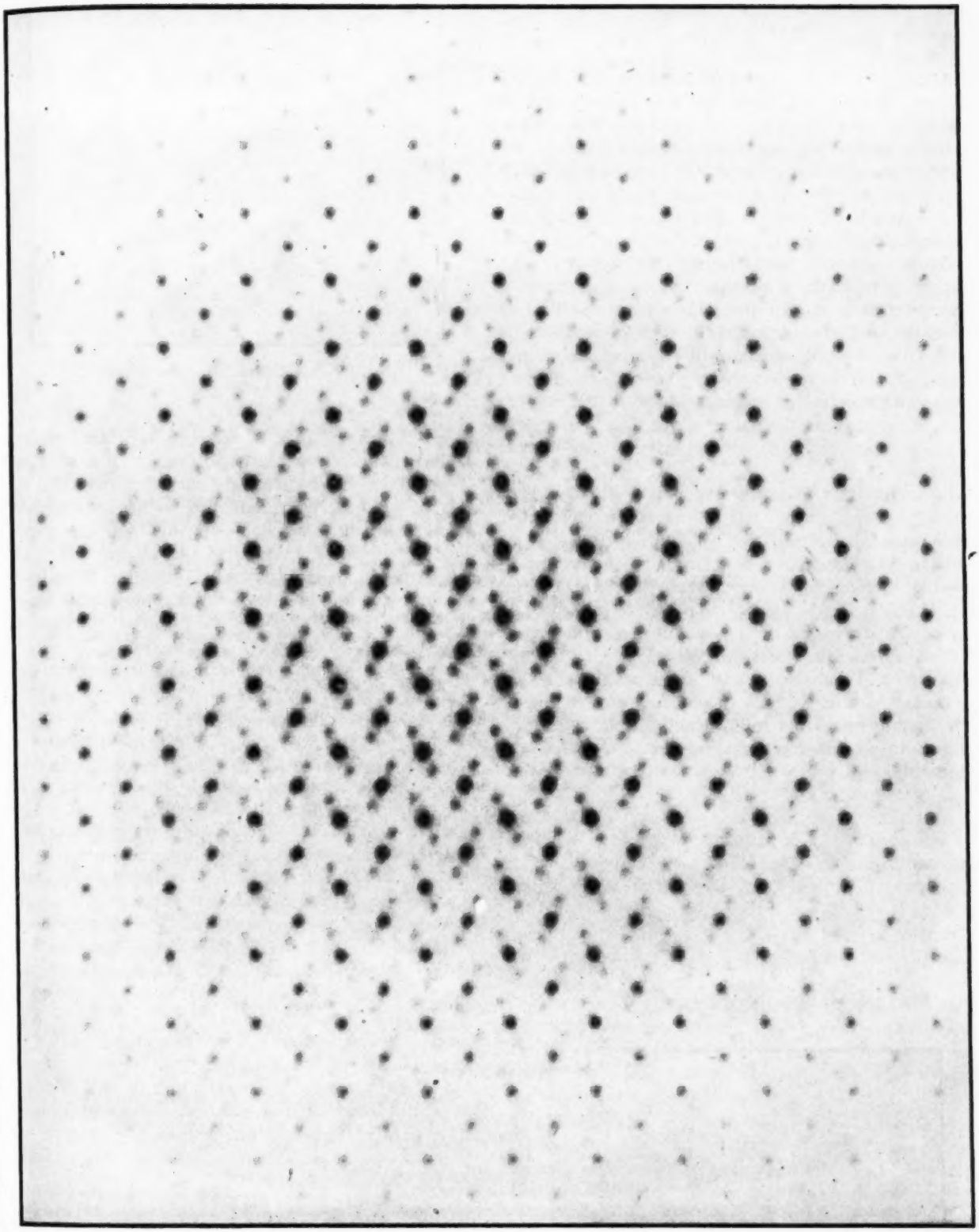


FIG. 13. Image of the atoms in *marcasite*, FeS_2, magnified 2.6×10^7 diameters. The larger dark circular areas are iron atoms, with 26 electrons each, while the fainter darkened areas are sulfur atoms with 16 electrons each.

mica is carried out about its optical direction Y, which minimizes the double refraction of the mica.

For imaging projections of crystal structures without centrosymmetry, each spectrum has, in general, its own different phase. To impress all these different phases that the phase-shifter block must have, the mica must be mounted so that its angle with the optical axis of the system can be varied. Figure 12 shows the design of such phase shifters which has been found to be practical. Between 200 and 400 of such units are needed to properly image the average crystal structure projection.

Figure 11 showed the phase shifters in place on the grating appropriate for imaging the crystal structure of marcasite, FeS_2 as seen along the c axis, and Fig. 13 shows the final image produced by the two-wave-length microscope system, utilizing this grating. The iron atoms, with 26 electrons each, are imaged as the heavier spots, while the sulfur atoms, with only 16 electrons each, are imaged as the lighter spots. The photograph is an accurate electron density map of this crystal structure.

To produce this image, the 150 individual $hk0$ spectra within recording range of the 30° precession photograph made with Mo$K\alpha$-radiation were utilized. A few of these spectra had amplitudes of zero. Of the entire 150 possible spectra, 32 required phase shifts of π, and Fig. 11 shows the 32 corresponding phase shifters. The test of the success of the phase shifting is the reproduction of the sulfur atoms, which are imaged as the less dense spots. To image the iron atom positions in this particular pattern requires the use of positive phases only, but to image the sulfur atom positions properly requires that the correct combination of negative and positive phases has been impressed on the grating.

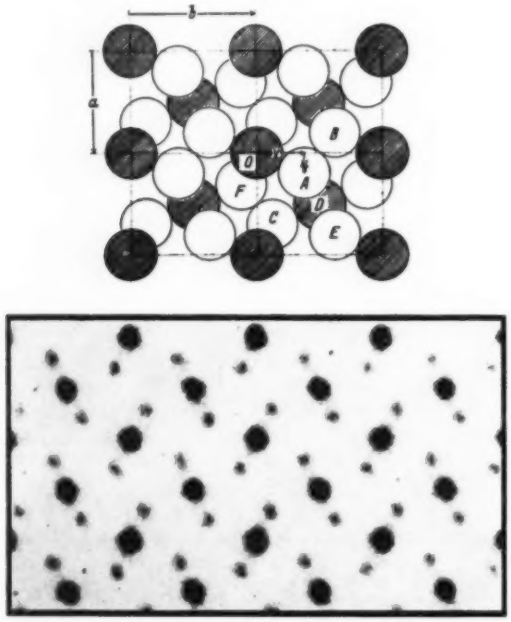


FIG. 14. Below: Image of the atoms in *marcasite*, FeS_2. Above: Diagrammatic representation of the crystal structure of *marcasite*.

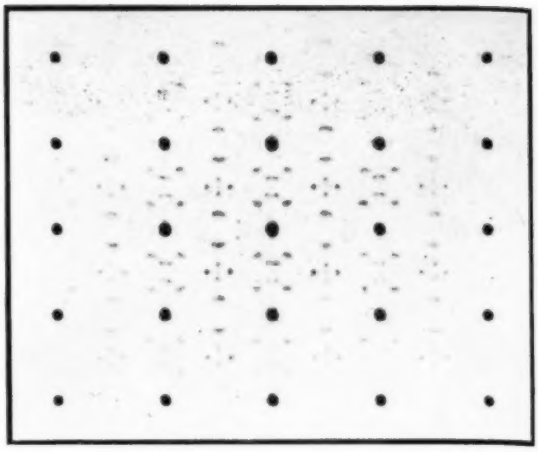


FIG. 15. Photographic Patterson synthesis, $P(xy)$ for *berthierite*, $FeSb_2S_4$.

Figure 14 shows part of Fig. 13 compared with a diagrammatic representation of the crystal structure of marcasite taken from an earlier paper.[7] In the diagrammatic representation, the shadings of the circles do not represent electron density, and their sizes merely illustrate the packing radii of the atoms, which are not related in a simple way to the electron densities in the atoms. With this proviso, the photographic reproduction of the atoms is in striking agreement with the published crystal structure of marcasite.

A word of caution is required. The development just described *does not* mean that now any crystal structure can be photographed. It *does* mean that this can be done provided that the phases which must be set with the mica phase shifters is known. These are the phases of the F's of x-ray diffraction, and they cannot be determined experimentally, although the magnitudes of the F's can. To know the phases requires, in effect, that the locations of the atoms are known. But when the phases are not known, which is the state of affairs as a new crystal structure is first examined, the apparatus described above automatically gives the Patterson synthesis[8,9] of the crystal, a fact which was noted[6] in 1939. An example of such a synthesis is shown in Fig. 15. The Patterson synthesis is related to the true image of the crystal structure in the following way: If one views the crystal structure from a fixed point, he sees a true image of the crystal structure. But if one occupies each of the n atoms of the structure in turn, viewing the crystal structure from each one, and then superposes all n parallel images, the composite image is the Patterson synthesis. Theoretical progress has recently

[7] M. J. Buerger, "Interatomic distances in marcasite and notes on the bonding in crystals of löllingite, arsenopyrite, and marcasite types," Zeits. f. Krist. (A)**97**, 504–513 (1937).

[8] A. L. Patterson, "A Fourier series method for the determination of the components of interatomic distances in crystals," Phys. Rev. **46**, 372–376 (1934).

[9] A. L. Patterson, "A direct method for the determination of the components of interatomic distances in crystals," Zeits. f. Krist. (A), 517–542 (1935).

been made in the direction of eliminating from the Patterson synthesis all crystal structure images but one.[10] When this theory and practice is perfected, it will become possible to set the phases for producing the true image by utilizing the first Patterson image made without phases.[11]

The writer is indebted to the National Research Council Committee on Scientific Aids to Learning for a grant in 1939 to cover the construction of the instrument described in this paper. He is also personally grateful to Drs. Vannevar Bush and Warren Mead for their encouragement in the development. The undertaking was greatly aided by the assistance of Mr. John Tyler, who superintended the engineering of the main optical system, turning ideas into realities. The gratings, phase shifters, and pinholes are the expert handiwork of Mr. John Solo, and the writer is also indebted to a graduate student, Mr. Jay W. Lathrop, for examining and calibrating suitable mica for the phase shifters used to make Figs. 13 and 14.

[10] M. J. Buerger, "Vector sets," Acta. Crystal. **3**, 87–97 (1950).
[11] M. J. Buerger, "Some new functions of interest in x-ray crystallography," Proc. Nat. Acad. Sci. **36**, 376–382 (1950).

A Model of Hurricane Formation*

HERBERT RIEHL
University of Chicago, Chicago, Illinois
(Received April 10, 1950)

At first, the description of the energy cycle of the mature tropical storm is amplified on the basis of recent upper air observations. Air particles passing through such a storm at first undergo isothermal expansion as they move toward a center. Then they ascend with release of condensation. At high levels they move outward and mix with the environment giving off heat to the surrounding colder air. Several requisites for maintenance of the observed temperature field are stated.

After a discussion of previous theories of hurricane formation the proposed model is described. The initial intensification of the wind field is brought about by mass divergence at high levels that imposes a pressure reduction on the surface layers. This divergence is the result of interaction between the large-scale disturbances of the upper air inside and outside the tropics. A solenoidal circulation is initiated that acts in the kinetic energy producing sense. But this circulation contains an internal mechanism for its own destruction and is maintained only under certain special conditions which are stated.

SINCE the early 1940's, numerous observations of pressure, temperature and wind have been taken in the atmosphere in the tropics at heights up to 20 km. These make possible a re-examination of the views held concerning the formation of tropical storms. Such a re-examination is the object of this report. New observational data are not offered. For these, the reader is referred to several of the references listed.[1–3]

THERMAL STRUCTURE OF FULLY DEVELOPED STORM

It is of advantage to begin with comments on the hurricane circulation after it has grown to full intensity. In some ways, tropical storms are those disturbances within the atmosphere whose maintenance is easiest to understand at present. Figure 1 is a representative sketch of pressure, temperature and circulation in a vertical section across a well-developed hurricane or typhoon outside the "eye." This figure shows a simple heat engine. Near the surface, air moves from high toward low pressure, converges and ascends. During the ascent, latent heat of condensation is liberated. In consequence the air near the center is warmer than its surroundings through a deep layer and the vertical spacing between the surfaces of constant pressure is larger. The pressure gradient force, directed toward the center, decreases upward and eventually reverses. Air particles moving upward near the center will be accelerated outward throughout the layer of decreasing pressure gradient, if their angular momentum is conserved.[4]

Figure 2 shows the vertical structure of the air surrounding tropical storms and a characteristic curve of the structure inside the rain area on a thermodynamic chart. Curve (b) is the same as that obtained by raising an air particle dry-adiabatically from sea surface to condensation level, and then moist-adiabatically to the 200-mb level. *It follows that the air mass encountered in the rain area below 200 mb consists entirely of particles that have ascended from the lowest levels.* Entrainment of mass from the surroundings does not take place, since such entrainment would produce upper temperatures considerably colder than those of curve (b). The air that

* A report on research conducted under contract between the ONR and the University of Chicago.
[1] H. Riehl, "On the formation of West Atlantic hurricanes," Dept. Meteorolog. Univ. Chicago, Misc. Rep., No. 24, (1948) 1–64.
[2] H. Riehl, "On the formation of typhoons," J. Meteorolog. **5**, (1948) 247–264.
[3] H. Riehl and N. M. Burgner, "Further studies of movement and formation of hurricanes," Bull. Am. Meteorolog. Soc. **31** (No. 7), (1950).
[4] C. S. Durst and R. C. Sutcliffe, "The importance of vertical motion in the development of tropical revolving storms," Quart. J. Roy. Meteorolog. Soc. **64**, (1938) 75–91.

fills the body of a hurricane originally is spread out in a thin sheet along the surface of the earth. This sheet—in which potential temperature and moisture content are nearly constant[5] has a depth of 50–60 mb. If a storm has a radius of 100 km, and if it extends to the 200-mb level, the surface air must be drained off from an area with a diameter of about 800 km in order to produce the first formation of the air mass observed inside the storm. This is an appreciable value.

One aspect of the temperature field has not received the attention it deserves. Byers[6] pointed out that the air spiralling inward toward a center experiences considerable expansion because of the pressure reduction which may have values from about 30 mb up to 100 mb. The temperature therefore should drop at a rate of nearly 1°C per 10 mb pressure reduction on account of adiabatic expansion. *Such decreases never occur. On the contrary, surface temperatures are remarkably constant throughout the body of tropical storms. The air moving inward expands at constant temperature,* and the potential temperature must increase from outskirts toward center. Rapid transfer of heat from ocean to atmosphere makes the isothermal expansion possible. In view of the highly agitated state of the ocean in hurricanes the occurrence of such rapid transfer is easily understood. Computations in a number of representative cases, mostly taken from data published by Deppermann[7], have verified the potential temperature increase. They have shown, moreover, that the moisture content of the air

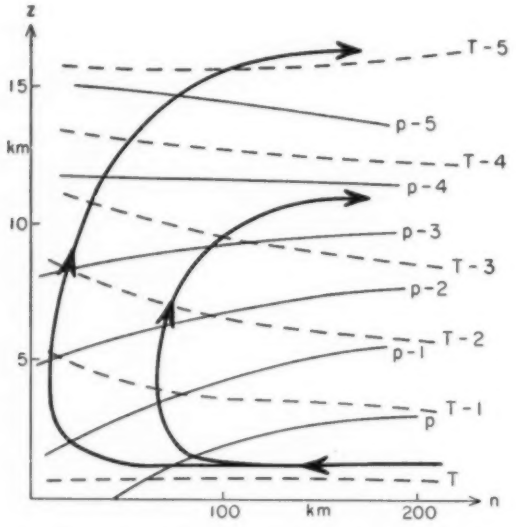


FIG. 1. Schematic outline of vertical circulation and distribution of temperature (T) and pressure (p) in rain area of a mature hurricane. The "z" axis points upward and the "n" axis outward across the storm from the eye boundary.

[5] Bunker, Haurwitz, Malkus, and Stommel, "Vertical distribution of temperature and humidity over the Caribbean Sea," Pap. Phys. Ocean. and Meteorolog. Mass. Inst. of Tech. and Woods Hole Ocean. Inst. 11, No. 1, (1949).

[6] H. R. Byers, *General Meteorology* (McGraw-Hill Book Co., Inc., New York, 1944).

[7] C. E. Depperman, *Some characteristics of Philippine typhoons* (Bureau of Printing, Manila, 1939).

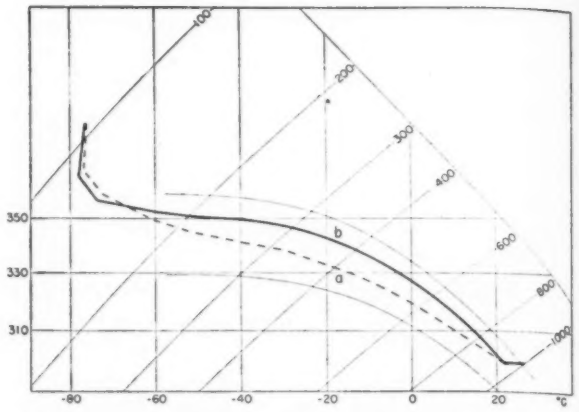


FIG. 2. Tephigram of (a) mean tropical atmosphere after Schacht (1946) and (b) sounding in rain area. Vertical lines are isotherms (°C), lines slanting from upper right to lower left isobars (mb) and horizontal lines are lines of equal potential temperature (dry-adiabats °A). Two moist adiabatic lines are entered on both sides of the soundings.

also becomes larger by 1 to 2 grams of water vapor per kg of air. This increase is also readily understood.

Figure 3 demonstrates the importance of this local energy source in tropical storms. Curve (b) represents the ascent of the surface air without local heat intake as in Fig. 2. Curve (a) shows an ascent from 960 mb under the assumption that the moisture content has risen by 1.5 g/kg and that the air has expanded isothermally to 960 mb. The difference between the two ascent curves is large. Comparison between curves (a) and (b) in Fig. 2 and curve (a) in Fig. 3 shows *that the area enclosed on the thermodynamic chart between the normal surroundings and the interior of a storm doubles in size if we allow for the local heat source.*

We must now turn to the problem of maintenance of the temperature field sketched in Fig. 1. The streamlines shown in this figure indicate outward transport of warm air aloft which should raise temperatures in the outskirts and destroy the solenoid field. If we make use of the local heat addition just mentioned, it is possible to obtain a temperature field that is partly steady. Let the air ascending near the outskirts of a cyclone follow curve (b) of Fig. 3, and the air nearest the center curve (a). Values intermediate between these extremes will be found in the middle region. If the air following curve (a) ascends the largest distance and the air following curve (b) the least, as sketched in Fig. 1, then the horizontal temperature gradient inside the storm will be maintained in the low and middle troposphere.

Outside the boundary of the storm, the outward transport of very warm air should be observed within a radius of 500 miles and more soon after establishment of a hurricane because of the large mass transport involved. Actually, however, the extremely warm air is not noticed at such a distance from the center. A slow and gradual warming of the surroundings takes place and very high temperatures are never reached. This

permits only one conclusion, namely that the air flowing out at high levels mixes with the surrounding air masses, that have not passed through the circulation, and gives up heat to them. At the same time the mixture is carried away from the storm area by the currents aloft and replaced by cooler air from other portions of the globe. In this way the entire temperature field relative to the storm is maintained. Two things must be noted. The mixing must take place outside the core since, as shown during the discussion of Fig. 2, the air of the environment does not enter the center aloft. It is also necessary to have a mechanism that is able to carry the mixture far away from the storm area and permit other air to approach.

Three phases of the energy cycle of the air passing through the cyclonic circulation have now been described. They are: (1) isothermal expansion with moisture addition; (2) dry-adiabatic expansion to the condensation level, followed by moist-adiabatic expansion; (3) mixing with cooler air aloft, a process that takes place at nearly constant pressure.

Figure 4 illustrates these three steps. For the construction of this figure a sample of air has been chosen that ascends to 200 mb and mixes with the surroundings at that level. The reduction of temperature from $-47°C$ to $-52°C$ at 200 mb, as indicated, is plausible, if we consider that the surroundings normally have a temperature of $-55°C$ which is observed to rise to about $-52°C$ in the situation under discussion. It is readily apparent from Fig. 4 that only a small percentage of the heat released during expansion of the ascending air from 1000 mb to 200 mb is available for conversion into kinetic energy. Maximum conversion is obtained if the air descends dry-adiabatically (following the dashed horizontal line) to the level at which its temperature becomes equal to that of the interior of the storm. The upper area (a) in Fig. 4 is only 12 percent of the total area (a+b) enclosed by the hurricane sounding and the normal surroundings. This represents the maximum efficiency of the circulation which, however, is never realized. As stated before, temperatures comparable to those inside are not observed outside. Therefore only a fraction of the energy represented by area (a) actually is used to drive the storm. The vast bulk of the heat released is spread over the globe.

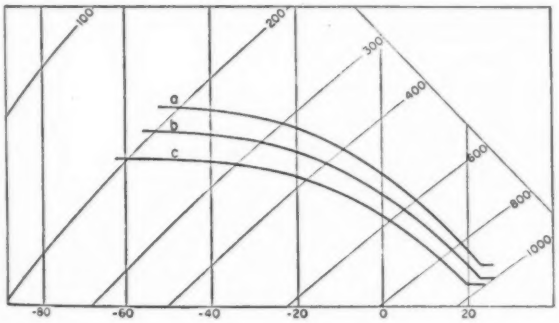


FIG. 3. Tephigram showing (a) ascent of air that has expanded isothermally to 960 mb at the sea surface with pickup of 1.5 g/kg of moisture; (b) same as sounding (b) of Fig. 2; (c) ascent of air with slightly polar characteristics (24°C, 16 g/kg).

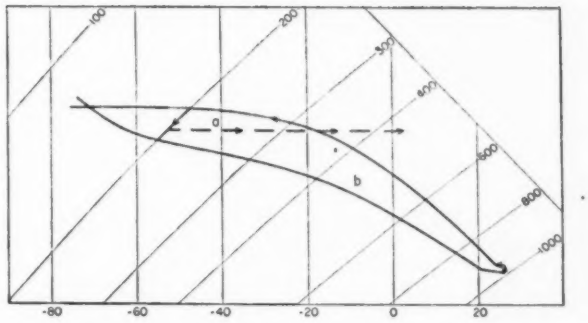


FIG. 4. Tephigram showing isothermal expansion of air at sea surface to 960 mb, dry-adiabatic ascent to condensation level, then moist-adiabatic ascent to 200 mb and horizontal mixing at 200 mb. Dashed horizontal line indicates descent of mixture outside the circulation. Lower curve is same as curve (a) of Fig. 2. The area (a+b) represents total heat gained by hurricane circulation, and area (a) represents maximum amount of energy available for conversion into kinetic energy.

Figure 4 shows clearly that a tropical storm and its surroundings cannot be considered as a closed system. It is a matter of weeks or more until the air that has risen in the storm can return to the surface under the influence of radiation. New air constantly enters the cyclone. In the course of the life of one storm, the air initially located over a large portion of a tropical ocean may pass through its circulation system.

Using the preceding discussion as a base, we are now in a position to examine critically several theories of hurricane formation that have been put forward.

PREVIOUS THEORIES OF HURRICANE FORMATION

Convectional Hypothesis

This is the oldest hypothesis and briefly states the following: an unusually large number of heavy rain squalls and thunderstorms develops for some reason over tropical ocean. These rain clouds then grow together and a cyclonic circulation develops at the ground due to convergence. Because of the heavy rainfall energy is available for an increase of the circulation (curve (b) of Fig. 2).

Most writers have discarded this hypothesis as insufficient for a number of years. But it still recurs in some textbooks. Here we shall mention only one objection: a mechanism for organized mass removal aloft is not specified, so that the area of heavy showers and its immediate surroundings will act almost as a closed system. After ascending the air will descend dry-adiabatically in the environment (following the horizontal lines of Fig. 2) and the environment will soon be warmer than the cloudy area. Any incipient circulation must dissipate. Actually, most heavy rainfalls of the tropics occur when the wind field is very weak or only slightly cyclonic. A sequence of events then takes place

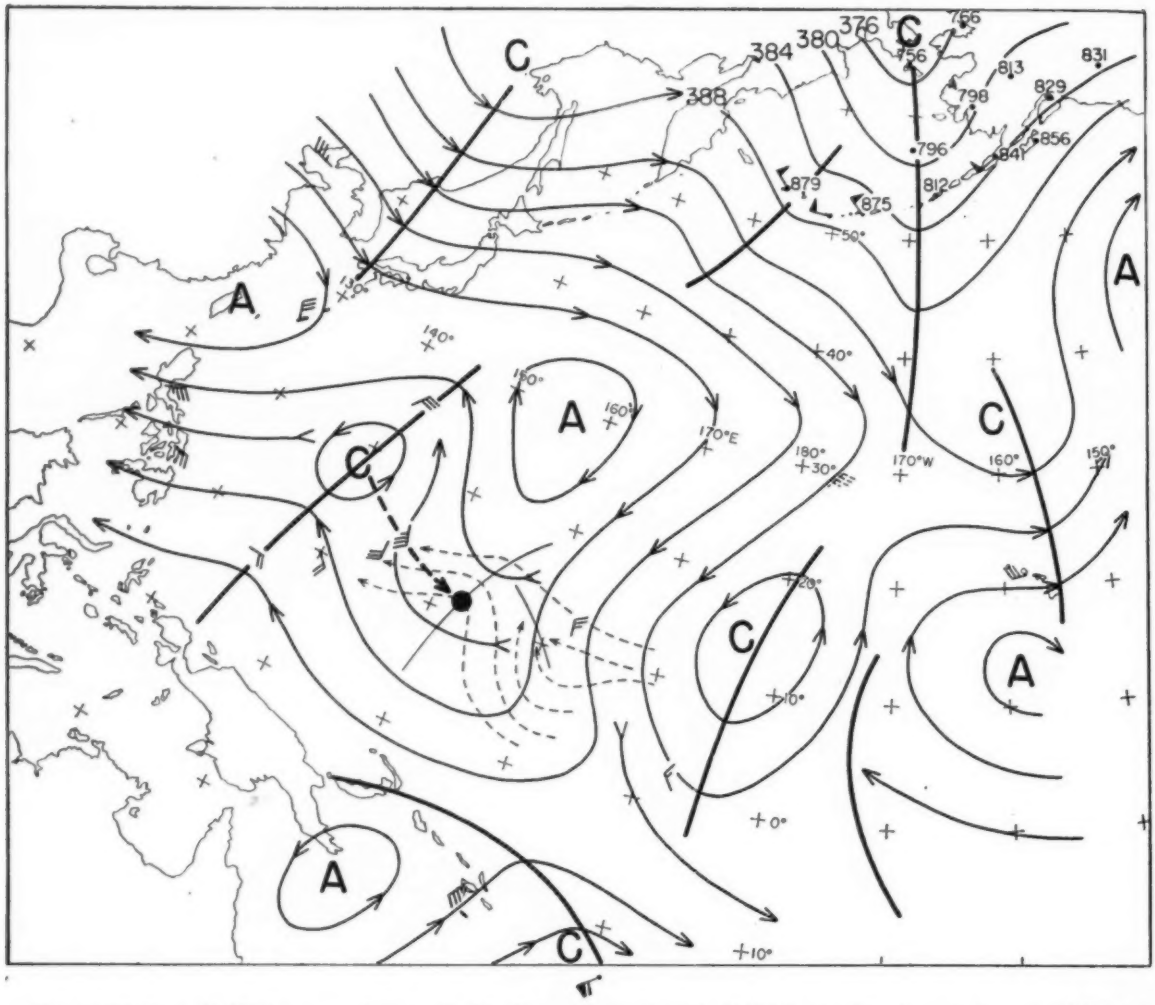


FIG. 5. Winds at 40,000 feet over the west Pacific Ocean, September 11, 1945.[2] In the Aleutians the lines are labeled as contours of the 200 mb surface (hundreds of feet). "*C*" stands for cyclonic circulation, "*A*" for anticyclonic circulation, and heavy solid lines indicate wave troughs and shear lines. In the wind reports a long barb denotes 10 m.p.h., a short barb five m.p.h. and a heavy triangular barb 50 m.p.h. Near the Marianas the 10,000-foot flow is shown by light dashed lines. Heavy dashed line denotes motion of incipient typhoon (heavy dot) relative to upper disturbance in which it had started during several preceding days.

which is fairly similar to that just outlined. Energy must be drawn into these circulations from an outside source in order to maintain them.

Frontal Hypothesis

About 1920, the Norwegian meteorologists formulated a theory of formation of shearing waves at density discontinuities (fronts). It has been since suggested that the zone of convergence, where the trades from northern and southern hemisphere meet, also is a boundary with a small density discontinuity along which shearing instability can develop. Actually, many storms, though by no means all of them, originate near this boundary. It is not possible, however, to accept the frontal hypothesis as an explanation.

TABLE I.

P	1	2	3	4
surface	26.0	26.0	26.0	24.0
900	20.0	20.5	22.5	18.0
700	8.0	11.0	13.5	8.0
500	−6.0	−3.0	0.5	−6.5
300	−33.0	−28.5	−23.5	−34.0
200	−55.0	−53.0	−47.0	−59.0

As in the case of the convection theory, a mechanism for outflow is not provided. There is another serious objection. According to the Norwegian concept, the denser air mass settles underneath the lighter mass, and this release of potential energy can lead to an increase of the kinetic energy. After some time (24–36 hours) the cyclone core at the surface consists entirely of cold air. Most of this cold air sinks in case of cyclones outside the tropics. In fundamental contrast, all air that enters a tropical cyclone near the surface ascends in the core. If a tropical cyclone core becomes filled with air that is even

slightly denser—and usually somewhat drier—than the normal tropical atmosphere, conditions become extremely unfavorable for the growth of a storm. Curve (c) of Fig. 3 illustrates this difficulty. This curve gives the ascent path of air that is very little cooler and drier ($T = 24°C$, specific humidity = 16 g/kg) than the air used for the construction of curve (b) ($T = 26°C$, specific humidity = 18 g/kg). A large temperature difference aloft results. If we compare curve (c) of Fig. 3 and curve (a) of Fig. 2 we see that no positive area is left on the thermodynamic chart. In fact, there is now a slight negative area. For comparison, the data of Figs. 2 and 3 are presented below in tabular form. Temperatures at certain isobaric surfaces are given for: (1) the average tropical atmosphere in the Caribbean; (2) the ascent of the surface air of average properties (26°C, 18 g/kg); (3) the ascent of the same air after isothermal motion to 960 mb with pickup of 1.5 g/kg of moisture; (4) the ascent of air with very slight polar characteristics (24°C, 16 g/kg).

Clearly, the entry of air with slightly polar characteristics must stop any incipient storm. It is a good forecasting rule that even intense hurricanes will weaken if they are affected by polar air while over a tropical ocean.

Hypothesis of "Dynamic Instability"

The foregoing theories are open to still another objection. Both take low level flow convergence as a starting point. Such convergence, however, must lead to pressure rises and not falls at the center, thus killing the beginning circulation. This is a well-known difficulty in cyclone theories. It is necessary to assume that the upper outflow somehow becomes sufficiently large to overbalance the surface inflow. To date, it has not been demonstrated how this is to be accomplished.

It is much more natural to suppose that at first a dynamic pressure fall occurs at the ground, produced by divergence at higher levels. Thereafter low level convergence sets in induced by the imposed pressure field. Sawyer[8] presented a hurricane theory based on this line of attack. He followed the work of Solberg[9] who showed that a symmetrical ring of air can become "dynamically unstable" under certain conditions. This can happen when the rotation of the air in space about the vertical axis (normal to the earth's surface at the place in question) is in the sense opposite to that of the rotation of the earth itself. Consider a ring of particles moving from west toward east along a latitude circle. If Solberg's criterion is satisfied, particles that are accelerated northward or southward will not tend to return to their starting latitude but move farther away from it. Under such circumstances dynamic pressure reductions can be produced.

Although Sawyer's attempt represents considerable progress, it is not conclusive for two main reasons. It is not indicated how the requisite sense of rotation is produced, nor is it shown how the unstable state is maintained. Even if "dynamic instability" exists in the region above a forming storm, the air evicted by this mechanism will sink in the surroundings. We then face the same reversal of the solenoid field which had made it impossible to accept the convectional hypothesis. No treatment exists as yet concerning the production of "dynamic instability." In middle latitudes, there frequently is a narrow zone in the high troposphere south of the maximum westerlies in which Solberg's criterion is satisfied.[10] Within the writer's experience, such a zone cannot be found in low latitudes, at least north of 15°N, in spite of the small value of the rotation of the earth itself (Coriolis parameter). Since data are scarce and scattered in the tropics, it is not easy to prove this assertion. Even at latitude 20°, however, a wind shear of 4 m sec.$^{-1}$/deg. lat., or a radius of flow curvature of 2° latitude at a windspeed of 10 m sec.$^{-1}$, are necessary in the limiting case when the rotation of the air particles relative to the earth is just equal in magnitude to the rotation of the earth itself. No plausible mechanism has been offered for the production of such large values over appreciable areas in the tropics prior to cyclone formation.

In summary, the theories just examined have one principal common drawback: they all consider only internal characteristics of the tropical atmosphere in the area of an initial disturbance. After some intensification this disturbance apparently is thought to be self-propagating. As we have seen, this is impossible. *At all times there must exist a connection between a hurricane and the surrounding atmosphere aloft over wide distances.* Sawyer's attempt is attractive in that it seeks to develop an initial field of upper divergence rather than a field of low level convergence. In the following, we shall retain this approach. But it is more plausible, and suggested by the observations, to specify an external rather than an

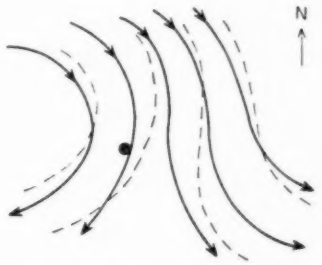


FIG. 6. Model of streamlines and contours (dashed lines) at 200 mb during deepening of a tropical disturbance (heavy dot) in northern hemisphere.

[8] J. S. Sawyer, Notes on the theory of tropical cyclones, Quart. J. Roy. Meteorolog. Soc. **73**, (1947) 101–126.

[9] H. Solberg, "Le mouvement d'inertie de l'atmosphère stable et son rôle dans la théorie des cyclones," Procès-verbaux de l'assoc. de meteorolog. Un. geod. geophys. Int., Edinbourg, (1936) 66–82.

[10] University of Chicago, Department of Meteorology, On the general circulation of the atmosphere in middle latitudes, Bull. Am. Meteorolog. Soc. **28**, (1947) 255–280.

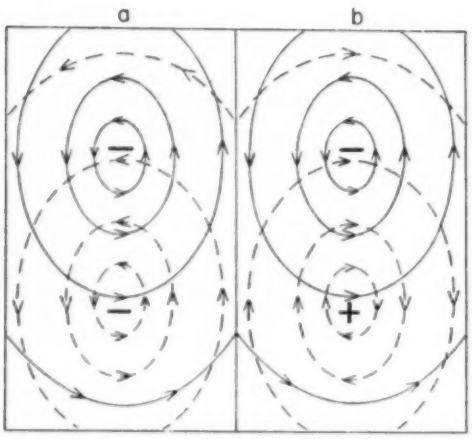


FIG. 7. Streamlines of perturbation centers at 200 mb (a) during in-phase and (b) during out-of-phase superposition of a trough in the westerlies on disturbances in the tropical vortex train aloft.

internal mechanism that can produce the requisite field of divergence.

PROPOSED MODEL OF HURRICANE FORMATION

Analysis of numerous charts at different levels over Atlantic and Pacific Ocean during the hurricane and typhoon season has brought out two factors that indicate the operation of external forces over an area within the tropics where a weak initial disturbance is present in the lower layers.[11]

(1) The trade winds of the low atmosphere disappear quickly with height near 500–400 mb (about 6 km). In the upper troposphere (300–150 mb, 9–15 km) we observe a train of large cyclonic and anticyclonic eddies that gradually drift westward[2] (Fig. 5). These eddies often move at a rate that differs from the displacement of the low level tropical disturbance. The latter, therefore, may be located under a different regime of high level winds at different times. Commonly, disturbances first develop in the air underneath an upper cyclone center. Deepening to typhoon strength, however, often takes place as the surface disturbance is being transferred from the influence of an upper cyclone to that of an upper anticyclone.

We shall choose a position of the deepening disturbance relative to the upper flow shown in Fig. 6 as a model. It is to be noted that the surface center and the air associated with it do not participate in the indicated motion of the current at 200 mb. This northerly current blows over the disturbed zone at rates that attain 50 m.p.h. and more.

(2) There is pronounced interaction at high levels between the tropical vortex train mentioned and the long waves in the westerlies outside the tropics. The low latitude eddies drift westward, and the troughs and ridges in the westerlies drift eastward. At times an upper cyclonic center of the tropics, for example, will be situated south of a trough in the westerlies, and at other times an anticyclonic center. Thus the connection between high level disturbances of middle and low latitudes will vary between in-phase and out-of-phase superposition. In Fig. 7 two types of superposition are sketched. Both disturbances have the same sense of rotation in Fig. 7a, but opposite rotation in Fig. 7b. The north-south components of motion are in the same sense in Fig. 7a, but the zonal component between the perturbation centers is opposite. In Fig. 7b, the reverse holds. Thus the amplitude of the composite streamlines and the intensity of the meridional flow will be increased by the superposition sketched in Fig. 7a, but the zonal speed in the central latitude band will decrease. Conversely, the meridional flow weakens in Fig. 7b and the streamline amplitude becomes flatter, while the zonal component in the central portion speeds up.

Without rigorous analysis we cannot know whether superposition produces net changes of kinetic energy. It is possible that the increase of north-south kinetic energy in case of Fig. 7a is obtained entirely through transformation of east-west into north-south kinetic energy. Our requirements are met if the speed of the northerly current shown in Fig. 6 increases in the region of the incipient tropical disturbance, irrespective of the net kinetic energy change over the whole system of middle and low latitude disturbances aloft. As just seen this occurs during in-phase superposition (cf. also Fig. 5), which therefore is requisite for the proposed model.

The Initial Pressure Fall

The effect of the superposition on the pressure field may be deduced from the equations of horizontal motion on the rotating earth

$$\dot{v}_n + v_s(kv + f - \dot{\alpha}) = -1/\rho(\partial p/\partial n), \tag{1}$$

$$\dot{v}_s - v_n(kv + f - \dot{\alpha}) = 0. \tag{2}$$

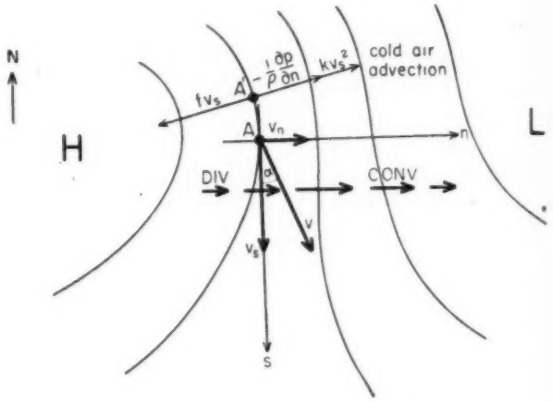


FIG. 8. Illustration of Eqs. (1), (2) at A and of Eq. (4) at A'. H denotes high pressure (resp. high elevation of 200 mb surface) and L low pressure (resp. low elevation of 200 mb surface).

[11] For a discussion of the formation of initial disturbances cf. reference 1.

Here, the coordinates s and n are chosen to lie parallel and normal (toward lower pressure) to the upper isobars (Fig. 8). v is the total motion, v_s the motion along and v_n the motion normal to the isobars (cross-stream circulation); f is the Coriolis parameter, p the pressure, ρ the density, k the trajectory curvature taken positive for cyclonic flow and α the angle between isobars and total motion (Fig. 8). The dots denote substantial differentiation with respect to time. Combination of (1) and (2) yields

$$\frac{1}{2}\frac{dv^2}{dt}=-\frac{v_n}{\rho}\frac{\partial p}{\partial n}, \tag{3}$$

which shows explicitly the well-known relation between acceleration, pressure gradient force and cross-stream flow. If the latter is zero prior to superposition, Eqs. (2) and (3) vanish and (1) becomes the gradient wind equation

$$kv_s^2+fv_s=-1/\rho(\partial p/\partial n), \tag{4}$$

where $v_s=v$ is the total motion. In case of anticyclonic flow the centrifugal force kv_s^2 and the pressure gradient force $-1/\rho(\partial p/\partial n)$ act to the left of the isobars looking downstream; the Coriolis force fv_s acts to the right (Fig. 8).

We now let the superposition take place, manifested by an increase of $-1/\rho(\partial p/\partial n)$. The Coriolis force fv_s no longer affords balance. A cross-isobar flow v_n toward lower pressure is indicated and also an increase of the total motion (kinetic energy). Figure 6 shows the streamlines of the total motion while the cross-stream flow is in progress.

On account of the motion normal to the isobars mass divergence takes place to the right of the northerly current looking downstream, and mass convergence to its left (Fig. 8). In the region of divergence the surface pressure falls because of the mass removal, and the initial surface disturbance begins to deepen. This is the desired result, and it is brought about entirely by a coincidence: presence of the surface depression under the northerly current at the time that in-phase superposition develops. Otherwise the superposition—an event observed in several areas of the globe on every day—has no particular consequences for reasons to appear later. Nor does the existence of a weak low-level depression in a position as sketched in Fig. 6 warrant a forecast of typhoon formation per se.

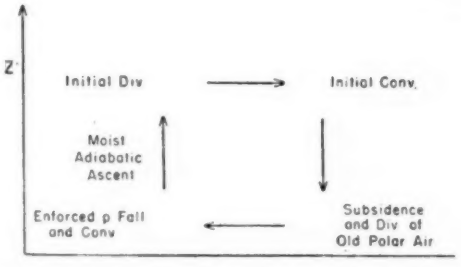


FIG. 9. Vertical cross-section along n-axis of Fig. 8. Model illustrates hypothetical development of vertical cross-stream circulation leading to deepening of the surface depression.

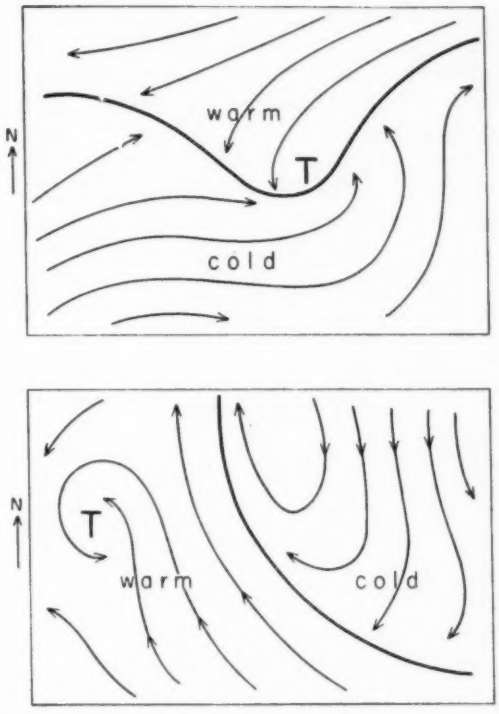


FIG. 10. Classical model of typhoon formation at "equatorial front" (above) and proposed revision (below) with cyclonic vortex situated entirely in the warm air.

Cross-Stream Circulation

The generation of the high level cross-stream flow and attendant field of convergence and divergence has several obvious consequences. Surface convergence develops under the region of upper divergence and, conversely, surface divergence develops under the region of upper convergence. It follows from continuity requirements that air must ascend along the right margin of the northerly current aloft and descend along its left margin (looking downstream). Figure 9 shows the complete cross-stream circulation in vertical cross-section along the n-axis, which points from high toward low pressure aloft. This circulation acts as a left-handed screw about the s-axis. The observations show that temperatures are higher on the right of this axis than on its left (Fig. 2). Therefore the warm air rises and the cold air sinks. The vertical circulation is in the kinetic energy producing sense.

We can now make use of another fact supplied by the observations. In the situation described a strong equatorward acceleration of the low level northeast trade frequently is noted along the eastern end of the upper northerly current and still farther east (500–1000 km from the tropical depression). Sometimes this acceleration is coupled with arrival of a polar air mass from the north.[3] Such polar air masses are gradually transformed

to tropical air during their equatorward passage. They are also observed to spread laterally and therefore sink. This sinking of old polar air adds to the descent in the eastern portion of the circulation cell shown in Fig. 9 and therefore acts to accelerate the whole cell. The polar air, of course, cannot invade the area of the tropical depression itself, if there is to be a storm. It bypasses this area going southward (Fig. 10). If, for some reason, the polar outbreak takes place too far to the west, the whole development ceases.

It is of interest that the model here proposed does find a role for old polar air, so prominently mentioned in many previous studies. This role differs radically from that formerly assigned to the polar air. Figure 10 illustrates the difference in viewpoint on the relative position of polar air and tropical disturbance. According to the present model, the sinking of the accelerated northeast trade provides a second mechanism that is suitable to remove air aloft from the area of the tropical depression and allow it to settle at a considerable distance. It is to be noted that this particular mechanism is not encountered in all instances. Among several alternatives, the formation of clockwise eddies in the low latitude trade stream has been mentioned.[2]

Maintenance of Cross-Stream Circulation

It is necessary now to examine the crucial problem of maintenance of the vertical circulation cell that has been initiated. There are two principal factors that act to restore equilibrium as soon as it is disturbed. One of these is the Coriolis force; the other the vertical stratification of the atmosphere. Except under special assumptions—which are often unrealistic—Eqs. (1, 2) do not yield a unique answer as there are too many variables. Consider, for instance, a situation near latitude 20° when initially $v_s = 10$ m sec.$^{-1}$ and $k = 10^{-6}$ m^{-1} (radius of curvature of 1000 km). Then $fv_s = 4 \times 10^{-2}$ cm sec.$^{-2}$ and $kv_s^2 = 10^{-2}$ cm sec.$^{-2}$. Let the windspeed v double in consequence of changes in the pressure gradient force Eq. (3). As $v_n \ll v_s$ in most situations of the kind here discussed, we can still estimate the magnitude of the forces in Eq. (1) by setting $v_s \sim v$. Both kv_s^2 and fv_s increase by 4×10^{-2} cm sec.$^{-2}$ provided k remains constant. In this case a true dynamic instability would have been achieved, especially as the term $-\dot{\alpha}v_s$, whose magnitude is hard to judge, acts in the same direction as the centrifugal force. The increase of restoring force is balanced by the increase of the centrifugal force. Thus the initial increment of the pressure gradient force remains unbalanced. As k is small, however, large percentual changes of this quantity are possible. Moreover, the cross-stream circulation affects the mass distribution and therewith the pressure gradient force. Only one definite statement is possible. A cross-stream circulation, once initiated, is much more easily maintained when the flow is anticyclonically curved than when it is cyclonically curved. In the latter case, fv_s and kv_s^2 act together as restoring forces.

The energy aspects yield a more conclusive answer. Let us return to curve (a) of Fig. 2 which serves as a representative sample of non-saturated air as encountered in the tropics. When such air is raised following the dry-adiabatic path indicated by the horizontal lines of Fig. 2, cooling results at all heights. When it sinks, warming takes place. Assuming that the vertical displacements sketched in Fig. 9 occur in non-saturated air, it follows that the western portion of the circulation cell will cool and that the eastern portion will become warmer. This acts to destroy the initial temperature gradient between eastern and western cell ends. It also acts to lower the pressure at high levels in the west and raise it in the east, reducing the increment in the pressure gradient force aloft that had been obtained through superposition. Clearly, the vertical circulation cell contains an internal mechanism for its own destruction under the circumstances described. The order of magnitude of vertical displacement requisite to stop the circulation is small. It may be estimated at 100 m.

It is quite probable that most vertical circulations actually come to an end in this way. If there is to be a hurricane, further necessary conditions must be introduced. At the eastern cell end, the temperature field relative to the incipient disturbance can be maintained if the warming due to subsidence is balanced by continuous advection of colder air. Circumstances are favorable for this to occur. The frequent equatorward movement of an old polar airmass in the low levels 500–1000 km east of a developing disturbance has already been mentioned. *At high levels, the superposition and attendant increase of streamline amplitude make possible prolonged displacement of cooler air from middle latitudes toward the tropics.* The data support this suggestion. Observations in a number of representative cases have shown that temperatures remained constant, or even decreased, in the eastern parts of the vertical circulation cells, in spite of subsidence.

In the region of ascent, temperature falls can be prevented by condensation. If the air that forms the tropical depression in the lower layers is saturated (area 10^3–10^4 sq. km), its ascent follows the moist rather than the dry-adiabatic path and no cooling will occur. *Thus it is essential to have such a saturated air mass available.* The saturation is produced by formation and dissipation of numerous cumulus clouds in the disturbed area. Moisture is funnelled upward by this mechanism and then spread laterally aloft. It is of importance to note that this process takes place mainly below six km. As stated initially, a rapid transition from trade wind to upper vortex regime is observed near that level. The lower air columns have only a slow drift relative to the surface depression. Whereas the air at 200 mb can pass over it in perhaps three hours, the low level air may remain in its orbit for days. Thus the moisture that is funnelled aloft,

is not carried away and gradual saturation below the level of wind reversal is made possible.

It is of much interest to note that such penetration of moisture is known to coincide with a gradual approach of the temperature lapse rate to the moist adiabatic. When saturation is complete, the vertical stratification of the atmosphere is nearly neutral and no longer unstable with respect to moist adiabatic displacements of particles. The local buoyancy force becomes relatively unimportant. This observation reveals another difference between dynamic and thermal theories of hurricane formation. The latter wish to obtain energy for the development of a circulation from vertical instability with respect to moist-adiabatic ascent. For practical purposes, this means that the atmosphere is unsaturated except for the rising parcels. As just seen, this is precisely the situation in which a large-scale vertical circulation cell will not be maintained. Of course, it would be best to operate with a field of upper divergence on a lower atmosphere that is saturated and has a temperature lapse rate in excess of the moist adiabatic. Since this cannot be obtained in reality, *an atmosphere as given by curve (b) of Fig. 2 represents optimum conditions for cyclone formation. Condensation energy is converted into kinetic energy not through local overturning of a vertically unstable atmosphere, but through large-scale vertical circulations. It follows that any mechanism, applied to the lower atmosphere, that operates to establish curve (b) of Fig. 2 will augment the chances of hurricane development.*

Conclusion

The foregoing pages have attempted to determine the reasons why tropical storms form under certain circumstances. In particular, we have sought to establish a mechanism for the removal of air over a cyclogenetic zone at high levels. This led to an inversion of previous attacks on the problem. Former investigators at first tried to find reasons for an increase of the cyclonic circulation from low-level considerations and then looked for ways and means to remove the air ascending in the vortex core. Such an approach restricts the analysis to internal factors—latent heat of condensation and internal structure of the field of motion.

In this report, the mechanism for outflow is established prior to deepening, and the cyclone is considered to deepen in response to upper divergence. A total of three energy sources has been applied to this end. The in-phase superposition of high and low latitude wave trains in the upper troposphere directly intensifies the speed of the high level northerly current that has been analyzed under the assumption that the balance of forces within it is the decisive factor. Subsidence of a southward moving mass of polar air provides a second energy source underneath the left-hand portion of this current, and the latent heat of condensation acts as third source on its right.

As yet, atmospheric processes cannot be measured directly but must be inferred from the three-dimensional structure of the atmosphere and its changes with time. These processes are not tied uniquely to the configurations of upper or lower flow pattern. We know, for instance, that it is not common to have ascent with high level divergence in the eastern portions of dynamic highs. Most of the time convergence and descent prevail aloft in these areas. Again, hurricanes have been observed to deepen when passing underneath upper troughs. But the shape of these troughs and the temperature distribution within them are not unique. Streamline and/or trajectory curvature can be cyclonic or anticyclonic.

Certain basic features of the *processes* that lead to cyclone formation must be alike if the same result is to be achieved at the surface. This study has attempted to uncover some of these processes. It is a task for the future to analyze other types of flow pattern configuration that lead to cyclogenesis.

ACKNOWLEDGMENT

The writer is greatly indebted to Dr. H. R. Byers, who took much time in helping to organize the material of this report. Drs. J. Bjerknes and E. Palmén discussed several aspects of the problems with the writer, and some phases of the study were undertaken following these discussions. Acknowledgment also is due to Dr. T. C. Yeh and Mr. N. La Seur, who participated in the analysis of many aspects of the topics treated.

Nuclei of Strain in the Semi-Infinite Solid

RAYMOND D. MINDLIN AND DAVID H. CHENG
Columbia University, New York, New York
(Received April 28, 1950)

Stress functions are given for forty nuclei of strain in the semi-infinite elastic solid. The corresponding stresses have the property that the traction vanishes across the plane boundary of the solid. The nuclei are derived, by processes of superposition, differentiation and integration, from the solution of the elasticity equations for the single force in the interior of the semi-infinite solid.

I. INTRODUCTION

THERE is an interesting group of solutions of the equations of the linear theory of elasticity known as nuclei of strain. For the isotropic solid of indefinite extent the fundamental solution of this type is that of Kelvin[1] for a single force applied at a point. By differentiation of this solution a family of additional nuclei can be obtained, examples of which are[1] the double force and the double force with moment. These, in turn, can be combined to form nuclei such as the center of dilatation and the center of rotation. By further differentiation, various types of doublets and multiplets are obtained. Alternatively, the preceding solutions may be integrated to form line distributions, such as[1] semi-infinite lines of centers of dilatation, or doublets, or double forces with moment, etc.

The solutions of many interesting problems may be expressed in terms of combinations of these nuclei or have been discovered by a process of superposition of nuclei. Examples are Lamé's problem[2] of the spherical container under internal and external pressure, Boussinesq's and Cerruti's problems[3] of the single force acting at the surface of a semi-infinite solid, Southwell's problem[2] of the spherical cavity in an infinite solid under simple tension, the analogous solution by Goodier[4] for the spherical inclusion and Mindlin's solution[5] for the single force in the semi-infinite solid.

The last named solution occupies the same position for the semi-infinite solid as Kelvin's solution does for the solid of indefinite extent. The nuclei derived from it, by differentiation and integration, all satisfy the condition of vanishing traction on a plane boundary so that, in contemplating the solution of new problems with at least one plane, traction-free surface, a portion of the boundary conditions is already satisfied.

Since the solution for the single force in the semi-infinite solid is itself formed from the superposition of eighteen nuclei derived from Kelvin's solution (six for each of the three components of the force), the calculation of nuclei for the semi-infinite solid is a formidable one. The possible usefulness of such nuclei is so great, however, that this task was undertaken and the results for forty nuclei are reported below.[6] The methods employed are the same as those described by Love[1] in connection with the calculation of nuclei in the solid of indefinite extent. The results are given in very condensed form by expressing them in terms of the Galerkin vector stress function. The displacement vector $\mathbf{u}$ is given in terms of the Galerkin vector $\mathbf{F}$ by

$$2G\mathbf{u}=2(1-\nu)\nabla^2\mathbf{F}-\boldsymbol{\nabla}\boldsymbol{\nabla}\cdot\mathbf{F}, \tag{1}$$

where G is the shear modulus, ν is Poisson's ratio and $\boldsymbol{\nabla}$, $\boldsymbol{\nabla}\cdot$ and ∇^2 are the gradient, divergence and Laplacian operators, respectively.

II. REDUCTION FROM THREE COMPONENTS TO ONE

Certain nuclei are obtained by the superposition of three others and the immediate result is a three-component Galerkin vector. This may be simplified, in some cases, by reduction to a one-component vector, say $\mathbf{F}'=\mathbf{k}Z'$, so that

$$2G\mathbf{u}=2(1-\nu)\mathbf{k}\nabla^2Z'-\boldsymbol{\nabla}(\partial Z'/\partial z). \tag{2}$$

To find the function $\mathbf{F}'$ corresponding to a given $\mathbf{F}$, (1) and (2) are equated to obtain the differential equation

$$2(1-\nu)\mathbf{k}\nabla^2Z'-\boldsymbol{\nabla}(\partial Z'/\partial z)=2(1-\nu)\nabla^2\mathbf{F}-\boldsymbol{\nabla}\boldsymbol{\nabla}\cdot\mathbf{F}, \tag{3}$$

which is to be solved for Z' in terms of $\mathbf{F}$.

Case 1: $\nabla^2\mathbf{F}=0$

Operating on (3) with $\boldsymbol{\nabla}\cdot$ and $\boldsymbol{\nabla}\times$ and using the condition $\nabla^2\mathbf{F}=0$, there results

$$\boldsymbol{\nabla}\nabla^2Z'=0$$

from which ∇^2Z' is a constant which may be taken as zero since, by (2), it contributes only a rigid body displacement. Hence (3) reduces to

$$\boldsymbol{\nabla}(\partial Z'/\partial z)=\boldsymbol{\nabla}\boldsymbol{\nabla}\cdot\mathbf{F} \tag{4}$$

the first integral of which is

$$\partial Z'/\partial z=\boldsymbol{\nabla}\cdot\mathbf{F}$$

[1] A. E. H. Love, *Mathematical Theory of Elasticity* (Cambridge University Press, London, 1927), fourth edition, Chapter VIII.
[2] S. Timoshenko, *Theory of Elasticity* (McGraw-Hill Book Company, Inc., New York, 1934), Chapter 11.
[3] See reference 1, Chapter X.
[4] J. N. Goodier, "Concentration of stress around spherical and cylindrical inclusions and flaws," J. App. Mech. **1**, 39–44 (1933).
[5] R. D. Mindlin, "Force at a point in the interior of a semi-infinite solid," Physics **7**, 195–202 (1936).
[6] An application of one of them is given in J. App. Phys. **21**, 931 (1950).

plus a constant which may be omitted since it does not contribute to the displacement. The second integral of (4) is

$$Z'=\int \nabla\cdot\mathbf{F}dz$$

plus a function of x and y which may be omitted since it does not contribute to the displacement.

Case 2: $\nabla\times\nabla\times\mathbf{F}=\nabla^2\nabla\cdot\mathbf{F}=0$

Since

$$\nabla\times\nabla\times\mathbf{F}=\nabla\nabla\cdot\mathbf{F}-\nabla^2\mathbf{F},$$

Eq. (3) reduces, in this case, to

$$2(1-\nu)\mathbf{k}\nabla^2Z'-\nabla(\partial Z'/\partial z)=(1-2\nu)\nabla\nabla\cdot\mathbf{F}. \quad (5)$$

Operating on (5) with $\nabla\cdot$ and $\nabla\times$ and using the condition $\nabla^2\nabla\cdot\mathbf{F}=0$, we find, as before, that $\nabla^2Z'=0$. Hence (5) reduces to

$$\nabla(\partial Z'/\partial z)=-(1-2\nu)\nabla\nabla\cdot\mathbf{F}$$

from which, by the procedure used in Case 1,

$$Z'=-(1-2\nu)\int\nabla\cdot\mathbf{F}dz.$$

III. HALF-SPACE NUCLEI IN TERMS OF GALERKIN VECTORS

The half-space considered is defined by $Z\geqslant 0$. Point nuclei are at $(0, 0, c)$ and one end of each line nucleus is at the same point. $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ are unit vectors in the x, y, z directions, respectively, and

$$R_1^2=x^2+y^2+(z-c)^2, \quad R_2^2=x^2+y^2+(z+c)^2.$$

A multiplying constant is omitted in each stress function.

A. Single Force

(1) Single force in x direction

$$\mathbf{i}\left\{R_1+R_2-\frac{2c^2}{R_2}+4(1-\nu)(1-2\nu)\times[(z+c)\log(R_2+z+c)-R_2]\right\}+\mathbf{k}\left[\frac{2cx}{R_2}+2(1-2\nu)x\log(R_2+z+c)\right].$$

(2) Single force in y direction

$$\mathbf{j}\left\{R_1+R_2-\frac{2c^2}{R_2}+4(1-\nu)(1-2\nu)\times[(z+c)\log(R_2+z+c)-R_2]\right\}+\mathbf{k}\left[\frac{2cy}{R_2}+2(1-2\nu)y\log(R_2+z+c)\right]$$

(3) Single force in z direction

$$\mathbf{k}\left\{R_1+[8\nu(1-\nu)-1]R_2+4(1-2\nu)[(1-\nu)z-\nu c]\log(R_2+z+c)-\frac{2cz}{R_2}\right\}.$$

B. Double Force

(1) Double force in x direction

$$\mathbf{i}x\left[\frac{1}{R_1}+\frac{1}{R_2}-\frac{4(1-\nu)(1-2\nu)}{R_2+z+c}+\frac{2c^2}{R_2^3}\right]+\mathbf{k}\left[\frac{2c}{R_2}-\frac{2cx^2}{R_2^3}+\frac{2(1-2\nu)x^2}{R_2(R_2+z+c)}+2(1-2\nu)\log(R_2+z+c)\right].$$

(2) Double force in y direction

$$\mathbf{j}y\left[\frac{1}{R_1}+\frac{1}{R_2}-\frac{4(1-\nu)(1-2\nu)}{R_2+z+c}+\frac{2c^2}{R_2^3}\right]+\mathbf{k}\left[\frac{2c}{R_2}+2(1-2\nu)\log(R_2+z+c)+\frac{2(1-2\nu)y^2}{R_2(R_2+z+c)}-\frac{2cy^2}{R_2^3}\right].$$

(3) Double force in z direction

$$\mathbf{k}\left[\frac{z-c}{R_1}-(1-4\nu)\frac{z-c}{R_2}-\frac{2cz(z+c)}{R_2^3}+4\nu(1-2\nu)\log(R_2+z+c)\right].$$

C. Double Force with Moment

(1) Double force in x direction with moment about y axis

$$\mathbf{i}\left[\frac{z-c}{R_1}-\frac{z-3c}{R_2}-\frac{2c^2(z+c)}{R_2^3}-4(1-\nu)(1-2\nu)\log(R_2+z+c)\right]+\mathbf{k}2x\left[\frac{c(z+c)}{R_2^3}-\frac{2(1-\nu)}{R_2}\right].$$

(2) Double force in x direction with moment about z axis

$$\mathbf{i}y\left[\frac{1}{R_1}+\frac{1}{R_2}+\frac{2c^2}{R_2^3}-\frac{4(1-\nu)(1-2\nu)}{R_2+z+c}\right]+\mathbf{k}2xy\left[\frac{1-2\nu}{R_2(R_2+z+c)}-\frac{c}{R_2^3}\right].$$

(3) Double force in y direction with moment about x axis

$$\mathbf{j}\left[\frac{z-c}{R_1}-\frac{z-3c}{R_2}-\frac{2c^2(z+c)}{R_2^3}-4(1-\nu)(1-2\nu)\log(R_2+z+c)\right]+\mathbf{k}2y\left[\frac{c(z+c)}{R_2^3}-\frac{2(1-\nu)}{R_2}\right].$$

(4) Double force in y direction with moment about z axis

$$\mathbf{j}x\left[\frac{1}{R_1}+\frac{1}{R_2}+\frac{2c^2}{R_2^3}-\frac{4(1-\nu)(1-2\nu)}{R_2+z+c}\right]+\mathbf{k}2xy\left[\frac{1-2\nu}{R_2(R_2+z+c)}-\frac{c}{R_2^3}\right].$$

(5) Double force in z direction with moment about x axis

$$\mathbf{k}y\left\{\frac{1}{R_1}+\frac{8\nu(1-\nu)-1}{R_2}+\frac{4(1-2\nu)[(1-\nu)z-\nu c]}{R_2(R_2+z+c)}+\frac{2cz}{R_2^3}\right\}.$$

(6) Double force in z direction with moment about y axis

$$\mathbf{k}x\left\{\frac{1}{R_1}+\frac{8\nu(1-\nu)-1}{R_2}+\frac{4(1-2\nu)[(1-\nu)z-\nu c]}{R_2(R_2+z+c)}+\frac{2cz}{R_2^3}\right\}.$$

D. Line of Double Forces with Moment

(1) Type C(1) along x axis from $x=0$ to $x=-\infty$

$$\mathbf{i}\left[(z-c)\log(R_1+x)-(z-3c)\log(R_2+x)-\frac{2c^2(z+c)x}{R_2(R_2^2-x^2)}\right]-\mathbf{k}\left[\frac{2c(z+c)}{R_2}+8\nu(1-\nu)R_2+4(1-\nu)(1-2\nu)(z+c)\log(R_2+z+c)\right].$$

(2) Type C(2) along x axis from $x=0$ to $x=-\infty$

$$\mathbf{i}y\left[\log(R_1+x)+\log(R_2+x)+\frac{2c^2x}{R_2(R_2^2-x^2)}\right]+\mathbf{k}y\left[2\nu(1-2\nu)\log(R_2+z+c)-2(1-\nu)(1-2\nu)\frac{z+c}{R_2+z+c}+\frac{2c}{R_2}\right].$$

(3) Type C(3) along y axis from $y=0$ to $y=-\infty$

$$\mathbf{j}\left[(z-c)\log(R_1+y)-(z-3c)\log(R_2+y)-\frac{2c^2(z+c)y}{R_2(R_2^2-y^2)}\right]-\mathbf{k}\left[\frac{2c(z+c)}{R_2}+8\nu(1-\nu)R_2+4(1-\nu)(1-2\nu)(z+c)\log(R_2+z+c)\right].$$

(4) Type C(4) along y axis from $y=0$ to $y=-\infty$

$$\mathbf{j}x\left[\log(R_1+y)+\log(R_2+y)+\frac{2c^2y}{R_2(R_2^2-y^2)}\right]+\mathbf{k}x\left[2\nu(1-2\nu)\log(R_2+z+c)-2(1-\nu)(1-2\nu)\frac{z+c}{R_2+z+c}+\frac{2c}{R_2}\right].$$

(5) Type C(5) along z axis from $z=c$ to $z=\infty$

$$\mathbf{k}y\left\{\log(R_1-z+c)+[8\nu(1-\nu)-1]\log(R_2+z+c)-\frac{4(1-2\nu)z}{R_2+z+c}-2\nu(1-2\nu)\left[\log(R_2+z+c)+\frac{R_2}{R_2+z+c}\right]-2z\left[\frac{1}{R_2}+\frac{z(z+c)}{R_2(x^2+y^2)}-\frac{z}{(x^2+y^2)}\right]\right\}.$$

(6) Type C(6) along z axis from $z=c$ to $z=\infty$

$$\mathbf{k}x\left\{\log(R_1-z+c)+[8\nu(1-\nu)-1]\log(R_2+z+c)-\frac{4(1-2\nu)z}{R_2+z+c}-2\nu(1-2\nu)\left[\log(R_2+z+c)+\frac{R_2}{R_2+z+c}\right]-2z\left[\frac{1}{R_2}+\frac{z(z+c)}{R_2(x^2+y^2)}-\frac{z}{(x^2+y^2)}\right]\right\}.$$

E. Center of Dilatation

(1)
$$\mathbf{i}[\log(R_1+x)+(1-4\nu)\log(R_2+x)]+\mathbf{k}\frac{2z}{R_2}.$$

(2)
$$\mathbf{j}[\log(R_1+y)+(1-4\nu)\log(R_2+y)]+\mathbf{k}\frac{2z}{R_2}.$$

(3)
$$\mathbf{k}\left[\log(R_1+z-c)+(1-4\nu)\log(R_2+z+c)+\frac{2z}{R_2}\right].$$

All three yield the same displacements and stresses.

F. Line of Centers of Dilatation

(1) Along x axis from $x=0$ to $x=-\infty$

$$\mathbf{i}\{x\log(R_1+x)-R_1+(1-4\nu)[x\log(R_2+x)-R_2]\}+\mathbf{k}2z\log(R_2+x).$$

(2) Along y axis from $y=0$ to $y=-\infty$

$$\mathbf{j}\{y\log(R_1+y)-R_1+(1-4\nu)[y\log(R_2+y)-R_2]\}+\mathbf{k}2z\log(R_2+y).$$

(3) Along z axis from $z=c$ to $z=\infty$

$$\mathbf{k}\{R_1-(z-c)\log(R_1+z-c)+(1-4\nu)\times[(z+c)\log(R_2+z+c)-R_2]+2z\log(R_2+z+c)\}.$$

G. Doublet

(1) Axis of doublet parallel to x axis

$$\mathbf{i}\left(\frac{1}{R_1}+\frac{1-4\nu}{R_2}\right)-\mathbf{k}\frac{2xz}{R_2^3}.$$

(2) Axis of doublet parallel to y axis

$$\mathbf{j}\left(\frac{1}{R_1}+\frac{1-4\nu}{R_2}\right)-\mathbf{k}\frac{2yz}{R_2^3}.$$

(3) Axis of doublet parallel to z axis

$$\mathbf{k}\left[\frac{1}{R_1}-\frac{1-4\nu}{R_2}+\frac{2z(z+c)}{R_2^3}\right].$$

H. Linearly Varying Line of Doublets with Strength Proportional to Distance from the Origin

(1) Type G(1) along y axis from $y=0$ to $y=-\infty$

$$\mathbf{i}\{R_1-y\log(R_1+y)+(1-4\nu)[R_2-y\log(R_2+y)]\}+\mathbf{k}\frac{2xz}{R_2+y}.$$

(2) Type G(1) along z axis from $z=c$ to $z=\infty$

$$\mathbf{i}\{R_1+z\log(R_1-z+c)+(1-4\nu)[R_2-z\log(R_2+z+c)]\}+\mathbf{k}2xz\left[\frac{1}{R_2}+\frac{z(z+c)}{R_2(x^2+y^2)}-\frac{z}{x^2+y^2}\right].$$

(3) Type G(2) along x axis from $x=0$ to $x=-\infty$

$$\mathbf{j}\{R_1-x\log(R_1+x)+(1-4\nu)[R_2-x\log(R_2+x)]\}+\mathbf{k}\frac{2yz}{R_2+x}.$$

(4) Type G(2) along z axis from $z=c$ to $z=\infty$

$$\mathbf{j}\{R_1+z\log(R_1-z+c)+(1-4\nu)[R_2-z\log(R_2+z+c)]\}+\mathbf{k}2yz\left[\frac{1}{R_2}+\frac{z(z+c)}{R_2(x^2+y^2)}-\frac{z}{x^2+y^2}\right].$$

(5) Type G(3) along x axis from $x=0$ to $x=-\infty$

$$\mathbf{k}\left\{R_1-(1-4\nu)R_2-x[\log(R_1+x)-(1-4\nu)\log(R_2+x)]-\frac{2z(z+c)}{R_2+x}\right\}.$$

(6) Type G(3) along y axis from $y=0$ to $y=-\infty$

$$\mathbf{k}\left\{R_1-(1-4\nu)R_2-y[\log(R_1+y)-(1-4\nu)\log(R_2+y)]-\frac{2z(z+c)}{R_2+y}\right\}.$$

I. Center of Rotation

(1) Center of rotation about x axis (Type C(5)—Type C(3))

$$\mathbf{j}\left[-\frac{z-c}{R_1}+\frac{z-3c}{R_2}+4(1-\nu)(1-2\nu)\log(R_2+z+c)+\frac{2c^2(z+c)}{R_2^3}\right]+\mathbf{k}y\left\{\frac{1}{R_1}+\frac{4\nu(1-2\nu)+3}{R_2}-\frac{2c^2}{R_2^3}+\frac{4(1-2\nu)[(1-\nu)z-\nu c]}{R_2(R_2+z+c)}\right\}.$$

(2) Center of rotation about y axis (Type C(6)—Type C(1))

$$\mathbf{i}\left[-\frac{z-c}{R_1}+\frac{z-3c}{R_2}+4(1-\nu)(1-2\nu)\log(R_2+z+c)+\frac{2c^2(z+c)}{R_2^3}\right]+\mathbf{k}x\left\{\frac{1}{R_1}+\frac{4\nu(1-2\nu)+3}{R_2}-\frac{2c^2}{R_2^3}+\frac{4(1-2\nu)[(1-\nu)z-\nu c]}{R_2(R_2+z+c)}\right\}.$$

(3) Center of rotation about z axis (Type C(4)—Type C(2))

$$(\mathbf{j}x-\mathbf{i}y)\left[\frac{1}{R_1}+\frac{1}{R_2}+\frac{2c^2}{R_2^3}-\frac{4(1-\nu)(1-2\nu)}{R_2+z+c}\right].$$

J. Line of Centers of Rotation

(1) Type I(1) along y axis from $y=0$ to $y=-\infty$

$$\mathbf{j}\Big[-(z-c)\log(R_1+y)+(z-3c)\log(R_2+y)+\frac{2c^2(z+c)y}{R_2(R_2^2-y^2)}\Big]+\mathbf{k}\Big\{R_1-[4(1-2\nu)^2-3]R_2+4(1-2\nu)[z+(1-2\nu)(z+c)]\log(R_2+z+c)+\frac{2c^2}{R_2}\Big\}.$$

(2) Type I(1) along z axis from $z=c$ to $z=\infty$

$$\mathbf{j}\Big\{R_1+R_2-\frac{2c^2}{R_2}+4(1-\nu)(1-2\nu)\times[(z+c)\log(R_2+z+c)-R_2]\Big\}+\mathbf{k}y\Big\{\log(R_1-z+c)+[2\nu(1-2\nu)+1]\log(R_2+z+c)+\frac{2c^2}{R_2(R_2+z+c)}-\frac{2[R_2(z-c)-z^2]}{x^2+y^2}+\frac{2(1-2\nu)}{R_2+z+c}[\nu R_2+2\nu(z+c)-2z]\Big\}.$$

(3) Type I(2) along x axis from $x=0$ to $x=-\infty$

$$\mathbf{i}\Big[-(z-c)\log(R_1+x)+(z-3c)\log(R_2+x)+\frac{2c^2(z+c)x}{R_2(R_2^2-x^2)}\Big]+\mathbf{k}\Big\{R_1-[4(1-2\nu)^2-3]R_2+4(1-2\nu)[z+(1-2\nu)(z+c)]\log(R_2+z+c)+\frac{2c^2}{R_2}\Big\}.$$

(4) Type I(2) along z axis from $z=c$ to $z=\infty$

$$\mathbf{i}\Big\{R_1+R_2-\frac{2c^2}{R_2}+4(1-\nu)(1-2\nu)\times[(z+c)\log(R_2+z+c)-R_2]\Big\}+\mathbf{k}x\Big\{\log(R_1-z+c)+[2\nu(1-2\nu)+1]\log(R_2+z+c)+\frac{2c^2}{R_2(R_2+z+c)}-\frac{2[R_2(z-c)-z^2]}{x^2+y^2}+\frac{2(1-2\nu)}{R_2+z+c}[\nu R_2+2\nu(z+c)-2z]\Big\}.$$

(5) Type I(3) along x axis from $x=0$ to $x=-\infty$

$$-\mathbf{i}y\Big[\log(R_1+x)+\log(R_2+x)+\frac{2c^2x}{R_2(R_2^2-x^2)}\Big]+\mathbf{j}\Big[R_1+R_2-\frac{2c^2}{R_2}\Big].$$

(6) Type I(3) along y axis from $y=0$ to $y=-\infty$

$$\mathbf{i}\Big[R_1+R_2-\frac{2c^2}{R_2}\Big]-\mathbf{j}x\Big[\log(R_1+y)+\log(R_2+y)+\frac{2c^2y}{R_2(R_2^2-y^2)}\Big].$$

IV. COMPONENTS OF DISPLACEMENT AND STRESS

The formulas for the components of displacement and stress corresponding to each of the nuclei listed in the preceding section are listed in Appendices I and II.[7]

[7] On file at the Department of Civil Engineering, Columbia University, New York 27, New York.

Thermoelastic Stress in the Semi-Infinite Solid

RAYMOND D. MINDLIN AND DAVID H. CHENG
Columbia University, New York, New York
(Received April 28, 1950)

The solution of the elasticity equations for the center of dilatation in the semi-infinite solid is introduced into Goodier's theory of thermoelastic stress. It is shown that the problem of potential to be solved is identical with that for the solid of indefinite extent. The results are applied to the case of an expanding (or contracting) spherical inclusion embedded in an elastic body near its surface.

I. INTRODUCTION

GOODIER[1] has shown that the elastic effect of a non-uniform distribution of temperature (T) in an infinite body is the same as that of a distribution of centers of dilatation of strength $-\beta/4\pi$ where

$$\beta=\alpha T(1+\nu)/(1-\nu) \tag{1}$$

and α and ν are the coefficient of linear thermal expansion and Poisson's ratio, respectively. Since α and T always appear together, what is said of a non-uniform distribution of temperature applies also to a space variation of the coefficient of linear thermal expansion. Goodier's theory has been applied to two problems of thermal stress in the infinite elastic body by Myklestad.[2]

If the body is not infinite the singularities within it give rise to tractions and displacements at its surface and the removal of these is a boundary-value problem in elasticity theory. However, Goodier noted that, if the solution of the elasticity equations for the center of dilatation is known for the bounded body, the thermoelastic problem is reduced, again, to an integration if the method of singularities is employed.

In this paper it is shown that, for the semi-infinite body with traction-free surface, the additional integral has the same form as that for the infinite body so that, if the latter is known, the calculation of displacement and stress reduces to a process of differentiation only. The result is then applied to a case of a spherical inclusion in the semi-infinite body. Use is made of the solution, given in an accompanying paper,[3] for the center of dilatation in the semi-infinite body.

II. THERMOELASTIC DISPLACEMENT

The Galerkin vector stress function for a center of dilatation of strength A at (ξ, η, ζ) in the infinite, isotropic, elastic solid is[4]

$$\mathbf{F}=-2AG\mathbf{k}\log(R_1'+z-\zeta),$$

where G is the modulus of rigidity, $\mathbf{k}$ is a unit vector in the z direction and

$$R_1'^2=(x-\xi)^2+(y-\eta)^2+(z-\zeta)^2.$$

Hence Goodier's nucleus of thermoelastic strain is given by

$$\mathbf{F}=(G\beta/2\pi)\mathbf{k}\log(R_1'+z-\zeta).$$

In an accompanying paper[3] it is shown that the corresponding nucleus at (ξ, η, ζ) in the semi-infinite body $z\geqslant 0$ (with the plane $z=0$ free of traction) is represented by

$$\mathbf{F}_0=(G\beta/2\pi)\mathbf{k}[\log(R_1'+z-\zeta)+(1-4\nu)\log(R_2'+z+\zeta)+2(z+\zeta)/R_2'-2\zeta/R_2'],$$

where

$$R_2'^2=(x-\xi)^2+(y-\eta)^2+(z+\zeta)^2.$$

Thus, for each center of dilatation at (ξ, η, ζ) in $z>0$, the plane $z=0$ is freed of traction by applying, at the image point $(\xi, \eta, -\zeta)$, a center of dilatation, a double force and a doublet, each of the proper strength.

The displacement $\mathbf{u}_0$ associated with $\mathbf{F}_0$ is calculated from

$$2G\mathbf{u}_0=2(1-\nu)\nabla^2\mathbf{F}_0-\boldsymbol{\nabla}\boldsymbol{\nabla}\cdot\mathbf{F}_0$$

and may be written in terms of operations on $1/R_1'$ and $1/R_2'$ as follows:

$$4\pi\mathbf{u}_0=-\beta\boldsymbol{\nabla}(1/R_1')-\beta\boldsymbol{\nabla}_2(1/R_2'),$$

where

$$\boldsymbol{\nabla}_2=(3-4\nu)\boldsymbol{\nabla}+2\boldsymbol{\nabla}z(\partial/\partial z)-4(1-\nu)\mathbf{k}\nabla^2 z$$

is a vector operator in which, for example, the scalar operator $\nabla^2 z$ signifies that the operand is first to be multiplied by z and then operated upon by ∇^2.

For a distribution $\beta(\xi, \eta, \zeta)$ in a region V_1 in $z>0$, the displacement is given by

$$\begin{aligned}4\pi\mathbf{u}&=4\pi\int_{V_1}\mathbf{u}_0 d\tau\\&=-\boldsymbol{\nabla}\int_{V_1}\frac{\beta}{R_1'}d\tau-\boldsymbol{\nabla}_2\int_{V_1}\frac{\beta}{R_2'}d\tau\\&=-\boldsymbol{\nabla}\int_{V_1}\frac{\beta}{R_1'}d\tau-\boldsymbol{\nabla}_2\int_{V_2}\frac{\beta}{R_1'}d\tau\\&=-\boldsymbol{\nabla}\varphi_1-\boldsymbol{\nabla}_2\varphi_2,\end{aligned} \tag{2}$$

[1] J. N. Goodier, "On the integration of the thermoelastic equations," Phil. Mag. **7** (No. 23), 1017–1032 (1937).
[2] N. O. Myklestad, "Two problems of thermal stress in the infinite solid," J. App. Mech. **9**, A-136-143 (1942).
[3] R. D. Mindlin and D. H. Cheng, "Nuclei of strain in the semi-infinite solid," J. App. Phys. **21**, 926 (1950).
[4] R. D. Mindlin, "Force at a point in the interior of a semi-infinite solid," Physics **7**, 195–202 (1936).

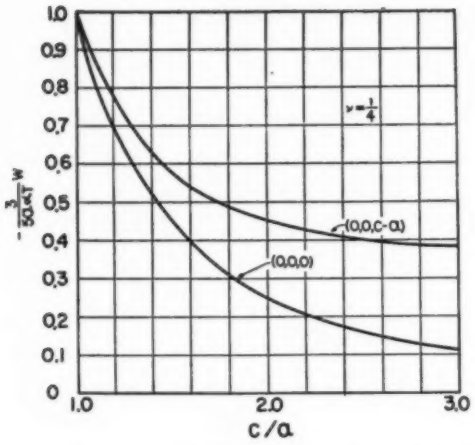


FIG. 1. Displacement normal to surface at points (0, 0, 0) and (0, 0, $c-a$).

where $d\tau=d\xi d\eta d\zeta$ and V_2 is the image of V_1 in the plane $z=0$. The volume integrals pass through ∇ and ∇_2 since the latter are x, y, z operators. The first integral is the potential (φ_1) of a distribution of matter of density β in V_1 as in Goodier's theory. The second integral (φ_2) is simply the reflection transformation of φ_1 in the plane $z=0$, so that the integration need not be repeated. For any distribution of β whose potential is known, the calculation of displacement in the semi-infinite body thus reduces to the simple process of applying the operators ∇ and ∇_2. Although φ_1 and φ_2 are mutual images, the corresponding displacements are not. This is due to the difference between ∇ and ∇_2.

III. APPLICATION TO SPHERICAL INCLUSION

Let V_1 be a sphere of radius a with center at (0, 0, c) where $a\leqslant c$. The radii to a field point (x, y, z) from (0, 0, c) and from its image (0, 0, $-c$) are denoted by $R_1=[x^2+y^2+(z-c)^2]^{\frac{1}{2}}$ and $R_2=[x^2+y^2+(z+c)^2]^{\frac{1}{2}}$ respectively. Let α be the excess of the coefficient of thermal expansion of the material within V_1 over that of the surrounding medium and let T be a uniform

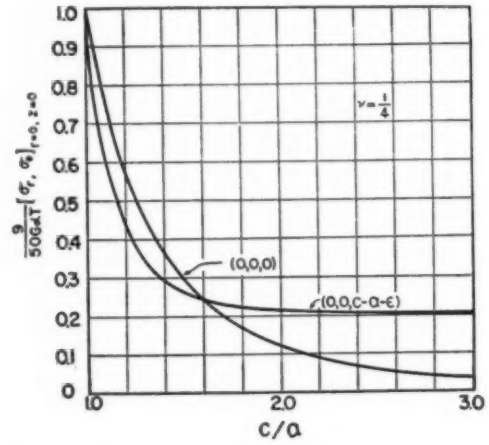


FIG. 2. σ_r and σ_θ at points (0, 0, 0), (0, 0, $c-a-\epsilon$).

temperature rise. Then β is given by (1) in $R_1<a$ and is zero in $z\geqslant 0$, $R_1>a$. The elastic constants of the two regions are assumed to be the same.

From the well-known results for the potential of a homogeneous sphere[5] we have, for exterior points ($z\geqslant 0$, $R_1>a$)

$$\varphi_1=4\pi a^3\beta/3R_1, \quad \varphi_2=4\pi a^3\beta/3R_2$$

and, for interior points ($R_1<a$)

$$\varphi_1=2\pi\beta(3a^2-R_1^2)/3, \quad \varphi_2=4\pi a^3\beta/3R_2.$$

Substituting in (2) and performing the operations ∇ and ∇_2, we have, for exterior points,

$$\mathbf{u}_e=\frac{a^3\beta}{3}\left\{\frac{\mathbf{R}_1}{R_1^3}+\frac{(3-4\nu)\mathbf{R}_2}{R_2^3}-\frac{6z(z+c)\mathbf{R}_2}{R_2^5}-\frac{2\mathbf{k}}{R_2^3}[(3-4\nu)(z+c)-z]\right\},$$

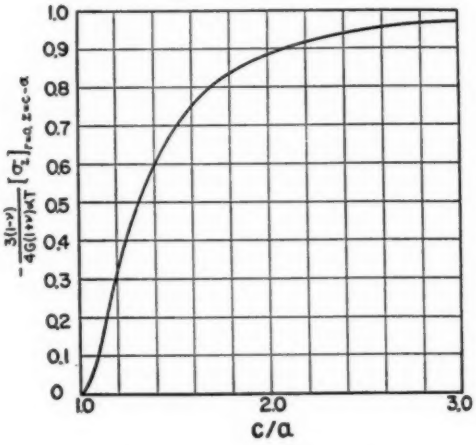


FIG. 3. σ_z at point (0, 0, $c-a$).

and, for interior points,

$$\mathbf{u}_i=\mathbf{u}_e+\frac{a^3\beta\mathbf{R}_1}{3}\left(\frac{1}{a^3}-\frac{1}{R_1^3}\right),$$

where $\mathbf{R}_1$ and $\mathbf{R}_2$ are the position vectors from (0, 0, c) and (0, 0, $-c$), respectively. It may be observed that, for exterior points, the displacement is the same as that for a single center of dilatation at the center of the sphere.

The stress is calculated from the displacement by means of the usual thermoelastic relation. In dyadic notation[6] this is

$$\Psi=\lambda\mathbf{I}[\nabla\cdot\mathbf{u}-\beta(1-\nu)/\nu]+G(\nabla\mathbf{u}+\mathbf{u}\nabla)$$

where Ψ is the stress dyadic, $\mathbf{I}$ is the idemfactor and

[5] W. D. MacMillan, *Theory of the Potential* (McGraw-Hill Book Company, Inc., New York, 1930), p. 40.
[6] C. E. Weatherburn, *Advanced Vector Analysis* (G. Bell and Sons, London, 1928).

$\lambda=2G\nu/(1-2\nu)$. In performing the calculations for interior points, β is given by (1) while for exterior points $\beta=0$. The cylindrical components are, for exterior points,

$$\sigma_r=\sigma_\theta-2G\beta a^3r^2\left[\frac{1}{R_1^5}+\frac{3-4\nu}{R_2^5}-\frac{10z(z+c)}{R_2^7}\right],$$

$$\sigma_\theta=\frac{2G\beta a^3}{3}\left[\frac{1}{R_1^3}+\frac{3-8\nu}{R_2^3}-\frac{6z(z+c)}{R_2^5}+\frac{12\nu(z+c)^2}{R_2^5}\right],$$

$$\sigma_z=\frac{2G\beta a^3}{3}\left[\frac{1}{R_1^3}-\frac{1}{R_2^3}-\frac{18z(z+c)}{R_2^5}+\frac{3(z+c)^2}{R_2^5}-\frac{3(z-c)^2}{R_1^5}+\frac{30z(z+c)^3}{R_2^7}\right],$$

$$\tau_{rz}=-2G\beta a^3r\left[\frac{z-c}{R_1^5}+\frac{3z+c}{R_2^5}-\frac{10z(z+c)^2}{R_2^7}\right],$$

$$\tau_{\theta r}=\tau_{z\theta}=0.$$

For interior points the formulas are the same except that in each of the normal components all the terms containing R_1 are replaced by $-2/a^3$, while the term containing R_1 in the shear component vanishes.

The magnitudes of the components of stress and displacement at the point on the sphere nearest the free surface and at the point on the free surface nearest the sphere are plotted in the Figs. 1, 2, and 3 as functions of the depth of the center of the sphere below the surface. It is interesting to observe that for an expanding sphere close to the surface the tensile stress in the surrounding medium can become $4(1+\nu)$ times as great as for the sphere in an infinite medium.

The Generalized Response of Linear Systems for Arbitrary Initial Conditions

JOHN R. MOORE
Aerophysics and Atomic Energy Division, North American Aviation, Inc. and Department of Engineering, University of California, Los Angeles, California

(Received May 1, 1950)

A method is presented for the solution of ordinary linear differential equations with constant coefficients. This method stays wholly in the time domain and obtains a general solution in a compact form including the effects of initial conditions. The differential equation as treated involves a driving function made up of the sum of terms involving the system input and its derivatives. The general solution is also given in terms of weighting functions operating on the input (in the driving function) and including outside the weighted integral, terms involving initial conditions of the input as well as the dependent variable.

INTRODUCTION

THIS paper presents a short method of computing the response of single input-single output linear systems with constant parameters to a large variety of driving functions. The result of this method is a general formula for the response expressed in terms of integrals of the input, the classical exponential transient terms and initial values of both the input and response functions together with their derivatives. The mathematical development is primarily in terms of functions in the time domain. The method and formulas are equally applicable to any linear differential equations with constant coefficients whether it be thought of in terms of a time dependent "input" and "response" or merely as a problem in mathematics.

SOLUTION OF THE DIFFERENTIAL EQUATION

The present paper takes the differential equation

$$G(p)\theta=KH(p)i \qquad (1)$$

[where θ is the dependent variable, i the input (function of variable t) $p\equiv d/dt$ and $G(p)$ and $H(p)$ are polynomials in p with constant coefficients] and obtains the general solution for θ, including the effect of initial conditions on θ and its derivatives.

The solutions may be made applicable to the case of multiple roots (two or more of the R_j equal) by standard limiting procedures.

The present method takes its cue from the classical method of successive substitution.* It was originally derived by the writer following a realization that the order of successive substitution is completely arbitrary and may be rearranged at will.

Starting with $G(p)$ in its factored form, n similar quantities, Θ_j are defined such that,

$$(p-R_j)\Theta_j\equiv G(p)\theta=KH(p)i. \qquad (2)$$

This is, apparently, equivalent to

$$\Theta_j=\frac{G(p)\theta}{p-R_j}=\frac{\prod_{i=1}^{n}(p-R_i)\theta}{p-R_j}=\frac{-\partial G(p)\theta}{\partial R_j}. \qquad (3)$$

* This is similar to the "Factorization Method" described in *Elementary Differential Equations*, by T. C. Fry (D. Van Nostrand Company, Inc., New York, 1929), seventh printing, pp. 182–186.

Solving (2) for Θ_j, n equations are obtained of the form

$$\Theta_j(t)=\Theta_j(0)e^{R_jt}+K\int_0^t e^{R_j(t-\tau)}H(p)i(\tau)d\tau. \quad (4)$$

Here

$$\Theta_j(0)=\frac{G(p)}{p-R_j}\theta\bigg|_{t=0}=\frac{-\partial G(p)}{\partial R_j}\theta\bigg|_{t=0}\equiv\frac{-\partial G(p_0)}{\partial R_j}\theta, \quad (5)$$

where

$$p_0{}^k\theta\equiv(d^k\theta/dt^k)\,|_{t=0}.$$

Since, by (2), each Θ_j contains θ and its derivatives through $p^{n-1}\theta$, it is possible to solve for θ from the n equations typified by (4) obtaining a linear combination of the various Θ_j. Although this could be done by the method of determinants, it is much simpler to proceed as follows.

Assume the linear combination is

$$\theta(t)=\sum_{v=1}^{n}k_v\Theta_v(t). \quad (6)$$

From Eq. (3),

$$\Theta_v(t)=\prod_{r=1}^{n}{}^{(v)}(p-R_r)\theta(t), \quad (7)$$

where the superscript on the product symbol denotes omission of the term for which $v=r$. Substitution of (7) into (6) gives:

$$\theta(t)\equiv\sum_{v=1}^{n}k_v\prod_{r=1}^{n}{}^{(v)}(p-R_r)\theta(t). \quad (8)$$

Equation (8) is an identity which must hold not only for all t but also for all functions $\theta(t)$. In particular, let $\theta(t)=e^{R_jt}$:

$$e^{R_jt}\equiv\sum_{v=1}^{n}k_v\prod_{r=1}^{n}{}^{(v)}(R_j-R_r)e^{R_jt}. \quad (9)$$

All of the products in Eq. (9) contain (R_j-R_j) as a factor except when $v=j$. Hence:

$$k_j=\frac{1}{\prod_{r=1}^{n}{}^{(j)}(R_j-R_r)}=\frac{1}{-\lim_{p\to R_j}[\partial G(p)/\partial R_j]}$$

$$\equiv\frac{1}{-[\partial G(R_j)/\partial R_j]}. \quad (10)$$

Substituting (10) and (4) into (6) gives the general formula:

$$\theta=\sum_{j=1}^{n}\left\{\frac{e^{R_jt}\dfrac{\partial G(p_0)}{\partial R_j}\theta-K\displaystyle\int_0^t e^{R_j(t-\tau)}H(p)i(\tau)d\tau}{[\partial G(R_j)/\partial R_j]}\right\}. \quad (11)$$

INTRODUCTION OF INPUT INITIAL CONDITIONS

Equation (11) contains $H(p)i$. This is often not as convenient to use as another form containing $H(R_j)i$ with initial conditions on i and its derivatives. To derive this form from (11), $H(p)$ is written as a summation, giving

$$\sum_{j=1}^{j=n}k_j\int_0^t e^{R_j(t-\tau)}H(p)i(\tau)d\tau$$

$$=\sum_{j=1}^{j=n}\sum_{\alpha=0}^{\alpha=m}\int_0^t k_je^{R_j(t-\tau)}h_\alpha p^\alpha i(\tau)d\tau. \quad (12)$$

Integrating each term by parts, gives elements of the form

$$\int_0^t e^{R_j(t-\tau)}p^\alpha i(\tau)d\tau$$

$$=(p^{\alpha-1}+R_jp^{\alpha-2}+\cdots+R_j{}^{\alpha-2}p+R_j{}^{\alpha-1})i(t)$$

$$-e^{R_jt}(p_0{}^{\alpha-1}+R_jp_0{}^{\alpha-2}+\cdots+R_j{}^{\alpha-2}p_0+R_j{}^{\alpha-1}p_0{}^0)i$$

$$+R_j{}^\alpha\int_0^t e^{R_j(t-\tau)}i(\tau)d\tau. \quad (13)$$

The form of (13) may be simplified by using the relation

$$(x-R_j)[x^{\alpha-1}+R_jx^{\alpha-2}$$

$$+\cdots+R_j{}^{\alpha-2}x+R_j{}^{\alpha-1}]=x^\alpha-R_j{}^\alpha \quad (14)$$

so that, symbolically

$$(p^{\alpha-1}+\cdots+R_j{}^{\alpha-1})i(t)\equiv[(p^\alpha-R_j{}^\alpha)/(p-R_j)]i(t) \quad (15)$$

and

$$e^{R_jt}(p_0{}^{\alpha-1}+\cdots+R_j{}^{\alpha-1}p_0{}^0)i$$

$$\equiv e^{R_jt}[(p_0{}^\alpha-R_j{}^\alpha p_0{}^0)/(p_0-R_j)]i. \quad (16)$$

Substitution of these equations into (12) gives

$$\sum_{j=1}^{j=n}k_j\int_0^t e^{R_j(t-\tau)}H(p)i(\tau)d\tau$$

$$=\sum_{j=1}^{j=n}k_jH(R_j)\int_0^t e^{R_j(t-\tau)}i(\tau)d\tau$$

$$+\sum_{j=1}^{j=n}k_j\sum_{\alpha=0}^{\alpha=m}h\alpha\left[\frac{p^\alpha-R_j{}^\alpha}{p-R_j}-e^{R_jt}\frac{(p_0{}^\alpha-R_j{}^\alpha p_0{}^0)}{p_0-R_j}\right]i$$

$$=\sum_{j=1}^{j=n}k_j\left\{H(R_j)\int_0^t e^{R_j(t-\tau)}i(\tau)d\tau\right.$$

$$\left.-e^{R_jt}\left[\frac{H(p_0)-H(R_jp_0{}^0)}{p_0-R_j}\right]i+\left[\frac{H(p)-H(R_j)}{p-R_j}\right]i\right\}. \quad (17)$$

This may be simplified by noting that the relations between the k_j and R_j make

$$\sum_{1}^{n} k_j R_j{}^{\alpha}=0, \quad (0\leq\alpha\leq n-2). \tag{18}$$

Thus, (17) becomes

$$\sum_{j=1}^{j=n} k_j \int_0^t e^{R_j(t-\tau)} H(p) i(\tau) d\tau$$

$$= \sum_{j=1}^{j=n} k_j \left\{ H(R_j) \int_0^t e^{R_j(t-\tau)} i(\tau) d\tau \right.$$

$$- e^{R_j t} \left[\frac{H(p_0) - H(R_j p_0{}^0)}{p_0 - R_j} \right] i$$

$$\left. + R_j{}^{n-1} \sum_{\beta=n}^{\beta=m} h_\beta \frac{(p^{\beta-n+1} - R_j{}^{\beta-n+1}) i}{p - R_j} \right\} \tag{19}$$

which makes the final solution for θ

$$\theta = -\sum_{1}^{n} \left[\frac{\partial G(R_j)}{\partial R_j} \right]^{-1}$$

$$\times \left\{ \frac{\{G(p_0)\theta - K[H(p_0) - H(R_j p_0{}^0)]i\}}{p_0 - R_j} \right\} e^{R_j t}$$

$$+ KH(R_j) \int_0^t e^{R_j(t-\tau)} i(\tau) d\tau$$

$$\left. + R_j{}^{n-1} \sum_{\beta=n}^{\beta=m} h_\beta \frac{(p^{\beta-n+1} - R_j{}^{\beta-n+1}) i}{p - R_j} \right\}, \tag{20}$$

where the last summation exists only if $H(p)$ is of the same order as, or higher order than $G(p)$.

CONCLUSION

The foregoing development attempts to carry out the general solution of the differential equation of an idealized linearized system including the often neglected effect of initial conditions both on the input and the response. The end formula is only slightly more difficult to remember than that of the Heaviside expansion theorem, yet covers a more general case. The method of solution is much simpler than the method of successive substitutions and yields more usable results for high order equations than the more usual type of Laplace transform solution.

BIBLIOGRAPHY

(1) G. Doetsch, *Theorie und Anwendung der Laplace-Transformation* (Dover Publications, New York, 1943, U. S. Alien Property Custodian), pp. 321–328.

(2) T. C. Fry, *Elementary Differential Equations* (D. Van Nostrand Company, Inc., New York, 1929), pp. 182–186.

(3) M. Gardner and J. L. Barnes, *Transients in Linear Systems* (John Wiley and Sons, Inc., New York, 1945).

(4) K. Wagner, *Operatorenrechnung* (Edwards Brothers, Inc., Michigan, 1944, U. S. Alien Property Custodian), pp. 94–98.

(5) Other standard books on differential equations.

Letters to the Editor

The Thermal Expansion Coefficient and the Melting Point of Cubic Elements

M. E. Straumanis
University of Missouri, School of Mines and Metallurgy, Department of Metallurgy, Rolla, Missouri
April 10, 1950

IN the Fig. 1 the linear expansion coefficient, as defined by $\alpha=(1dl)/(ldT)$, of cubic elements is plotted against the melting point T_F of the same elements.[1] The curve shows clearly that generally the linear expansion coefficient of cubic elements decreases with increasing melting point. There are no exceptions from this rule, only deviations from the curve. These deviations might be caused (1) by some differences in the structure of the cubic elements, or (2) by inaccurate determinations of the expansion coefficients at room temperature. Further work will show the nature of these deviations. The curve presented seems not to be a simple one, although the melting points of the 5 alkali elements fit well to the empirical equation $\alpha\sqrt[3]{T_F}=\text{const}$, T_F being expressed in °C.

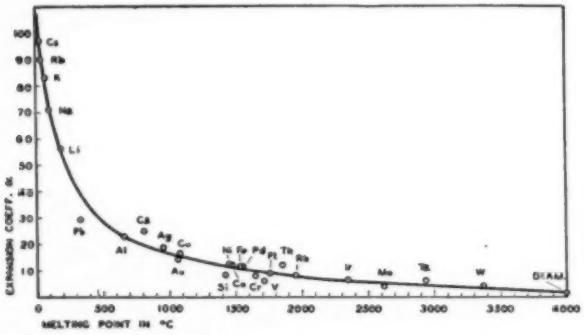


FIG. 1. Linear expansion coefficient $\alpha\times10^6$ of cubic elements at room temperature *versus* melting point of the same elements.

It might be that plotting the cubical expansion coefficient in the same manner as already shown, all at room temperature solid elements would be located on or near the curve.

No relations between melting point and expansion coefficient could be found in the literature, except some indications made by Grüneisen[2] and Lindemann.[3]

[1] The coefficients and melting points in Fig. 1 were obtained from the International Critical Tables, Vol. I, 103, from Hodgman's *Handbook of Chemistry and Physics* and from Landoldt-Börnstein, *Physikalischechemische Tabellen*.
[2] E. Grüneisen, Ann. d. Physik **39**, 297 (1912).
[3] F. A. Lindemann, Physik. Zeits. **11**, 609 (1910).

The Tracing and Interpretation of Asymmetrical Hysteresis Loops*

G. M. Ettinger**
Department of Electrical Engineering, New York University, New York, New York
May 3, 1950

IN the course of development of a hysteresis loop tracer with cathode-ray tube presentation (to be described elsewhere), the requirement arose to display the asymmetrical hysteresis loops[1] which are exhibited by magnetic materials subjected simultaneously to a.c. and d.c. magnetizing forces. Figures 1a to 1e show a series of loops which were obtained for "permalloy C" with increasing d.c. polarization. The polarizing current was produced by connecting in series with the magnetizing winding on the sample a metal rectifier shunted by a variable resistance.

To trace the curve of incremental permeability *vs.* d.c. polarizing field, the a.c. sensitivity of the (direct coupled) horizontal deflection amplifier in the hysteresis loop tracer was reduced to zero. The trace on the cathode-ray tube therefore changed to a straight line whose length was proportional to the peak flux density in the sample, while its displacement from the axis of B (shown dotted in Fig. 1) gave a measure of the d.c. magnetizing force. Such lines are superposed on their corresponding hysteresis loops in Fig. 1, while Fig. 2 shows a number of these lines photographed on the same film. Their envelope is the curve of $B_{a.c.}$ *vs.*

FIG. 1. Asymmetrica hysteresis loops.

FIG. 2. A.c. permeability *vs.* polarizing field for "permalloy C."

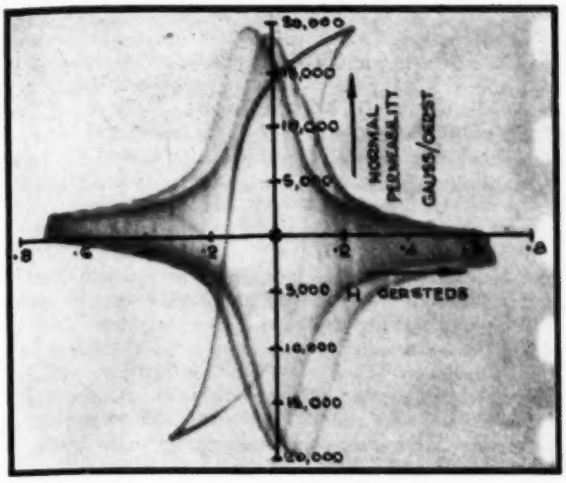


FIG. 3. Butterfly curve for "permalloy C."

$H_{d.c.}$. The peak value of the alternating component of magnetizing force was kept constant during the test, so that Fig. 2 represents, to a different scale, a portion of the $\mu_{a.c.}/H_{d.c.}$ or "butterfly" curve which is of importance in magnetic amplifier work. Other workers[2,3] have employed bridge or potentiometer methods to obtain this curve.

It is possible to observe the complete butterfly curve (from positive to negative saturation) on the screen of a long persistence cathode-ray tube if the potentiometer controlling the d.c. polarization is slowly driven to and fro by a reversing motor or a cam. Figure 3 is a photograph of the "butterfly curve" for "permalloy C," obtained by this method. A hysteresis loop is superposed for calibration purposes.

[1] L. F. Borg, J. Inst. Elec. Eng. (London) **96** (Part II), 317 (1949).
[2] G. W. Elmen, Nat. Bur. Stand. Tech. J. **15**, 124 (1936).
[3] B. Fertchak, Rev. Gen. Elec. (Paris) **54**, 79 (1945).
* This work was done at Standard Telecommunication Laboratories, Ltd., London, England.
** Department of Electrical Engineering, New York University.

Mechano Chemistry of the Dispersion of Mercury in Liquids in an Ultrasonic Field

EVELYN C. MARBOE AND W. A. WEYL
The Pennsylvania State College, School of Mineral Industries, State College, Pennsylvania
May 5, 1950

IN a previous publication[1] on the mechano-chemical properties of water it was pointed out that the strength of the binding forces between water and a metal may play a role in certain types of cavitation corrosion. It is the object of this report to describe some simple experiments which illustrate the role which the binding forces between metallic mercury and a liquid play on the dispersion of this metal in the liquid. For our problem, the use of liquid mercury offers the advantage that there can be no work-hardening of the metal, a factor which complicates the corrosion through cavitation for solid metals. Some parallel experiments were carried out with metallic gallium in order to make certain that the phenomena observed were not characteristic for mercury alone. All experiments were carried out with a quartz crystal type of high frequency generator. (Ultrason Model U-300 of the Televiso Products Company, Chicago, Illinois.)

Experiment No. 1. For a certain intensity of the ultrasonic field (0.2 kv, 20 ma at 25°C) the system, mercury-paraffin oil, was found to remain unaffected. Addition of stearic acid to the paraffin oil, however, caused dispersion of the metal. This experiment which confirms previous work[2] indicates clearly the importance of the forces acting between the metal and the hydrocarbon. A molecular film of Hg-stearate links together the incompatible media and, thus, makes dispersion possible and also stabilizes the emulsion by preventing the droplets from coalescing.

Experiment No. 2. The intensity of the ultrasonic field was adjusted (1 kv, 130 ma at 25°C) so that mercury in distilled water produced a slight turbidity after an exposure of five minutes. Under these conditions the water remained clear if it had been deaired by boiling. The effect of the air on the dispersion of the metal in water is explained as follows.

Oxidation of the mercury produces Hg^{+} and Hg^{2+}, both cations having a high polarizability. As has been pointed out in a previous report[3] ions of this type are adsorbed at a metal surface in a fashion which may be described as:

Metal—metallic bond—polarized cation—ionic bond—water.

By changing the metallic character of the metal surface into one resembling that of an ionic substance, the strong binding forces are provided between metal and water which are essential for the dispersion.

Experiment No. 3. The same experiment was repeated with distilled and deaired water. No turbidity was observed after five minutes exposure. Now a few drops of a saturated $HgCl_2$ solution were added in order to provide Hg^{2+} ions. Immediately a deep gray dispersion of Hg droplets in the diluted $HgCl_2$ solution began to form.

The connecting link between metal and water is the Hg^{2+} ion which becomes strongly polarized in the asymmetrical force field of the metal-water interface. Addition of KCl solution to the Hg^{2+} containing water changes the Hg^{2+} into a complex anion of the type $(HgCl_4)^{2-}$ and, as a result, mercury is no longer easily dispersed. Only the cation Hg^{2+}, but not the anion $(HgCl_4)^{2-}$, has the metallophilic properties essential for producing dispersion and stabilizing the emulsion. This experiment has its analog in the effect of KCl upon the solubility of metallic lead in fused lead chloride.

R. Lorenz[4] determined the influence of alkali chloride on the solubility of metallic lead in fused $PbCl_2$ at 610°C (Table I).

TABLE I. Solubility of metallic lead in the system $PbCl_2$—KCl.

$PbCl_2$	KCl	Solubility ×10,000
10 mole	0 mole	3.74%
9 mole	1 mole	2.27%
8 mole	2 mole	1.51%
7 mole	3 mole	0.64%
6 mole	4 mole	0.06%

The solubility of the metal in the fused salt is the result of the presence of Pb^{2+} ions which, in the asymmetrical force field of the Pb—$PbCl_2$ interface, are strongly deformed and thus are able to link the metal atoms with the ionic medium. Addition of KCl to the melt changes the $PbCl_2$ into K_2PbCl_4. The anion $(PbCl_4)^{2-}$ cannot act as a metallophilic group.

Experiment No. 4. Liquids containing polar molecules exert image forces upon the mercury and, as a result, increase the dispersability of the mercury in a liquid. Under conditions which did not lead to a dispersion of Hg metal in the non-polar carbon tetrachloride, the metal could be dispersed in the strongly polar nitrobenzene.

Summary. Dispersion of Hg in a liquid under the influence of ultrasonic vibrations requires the presence of strong forces between the metal and the medium. Strengthening the forces between the two incompatible media enhances dispersion and stabilizes the emulsion. This can be achieved in three ways:

1. Compound formation between the metal and a part of the liquid, e.g., addition of stearic acid to paraffin oil leading to a molecular film of the metal stearate.

2. Image forces induced in the metal surface by molecules with a permanent dipole, e.g., nitrobenzene.

3. Metallophilic groups such as strongly polarizable cations, e.g., Hg^{2+}. The mechanism of their action is known from several phenomena involving adhesion[5] (silver to glass by means of tin ions) and solution[6] (gold ruby glass requires the presence of metallophilic ions, such as Sn^{4+}, Bi^{3+}, Pb^{2+}).

[1] W. A. Weyl and E. C. Marboe, Research **2**, 19 (1949).
[2] L. N. Solov'eva, Colloid. J. (U.S.S.R.) **5**, 289–97 (1939).
[3] Enright, Marboe, and Weyl, ONR Tech. Report No. 1, Contract No. N6 onr 269, Task Order 8 NR 032–265, 11 pages (October, 1948).
[4] R. Lorenz and W. Eitel, *Pyrosole* (Academ. Verlags-ges., Leipzig, 1926), in particular pp. 54 and 55.
[5] W. A. Weyl, The Glass Industry **26** (12), 557 (1945).
[6] W. A. Weyl, J. Soc. Glass Tech., "Colors Produced by Metals. Part IV." **29**, 302–307 (1945).

The Barium Oxide on Tungsten Cathode Interface*

EUGENE B. HENSLEY AND JOHN H. AFFLECK
University of Missouri, Columbia, Missouri
May 17, 1950

IT is now recognized that the electrical behavior of oxide coated cathodes may be influenced either directly or indirectly by the presence of an interface compound formed by the reaction of the oxide coating with the base metal.[1] Tungsten and tungsten nickel alloys have been used as the base metal in experimental tubes and have found some commercial application in certain subminiature tubes. We have identified the interface compound for these cathode systems as being composed principally of barium tungstite, $BaWO_3$, when the coating is pure BaO. The corresponding tungstites are found when SrO or the solid solution (BaSr)O are used for the coating. An extensive search of the literature has failed to reveal any mention of the existence of these compounds; consequently particular care has been taken in their identification. Because of the small amounts of the interface compounds formed, x-ray diffraction methods of analysis have been used.

Cathodes were prepared by spraying $BaCO_3$ with an organic binder onto 10 mil tungsten wires which had been cleaned by electrolyzing in a KOH bath. The cathodes were then mounted in vacuum tubes and exhausted to a final pressure of 5×10^{-8} mm of Hg. The $BaCO_3$ was then reduced to the oxide by passing heater current through the tungsten wires and the tubes were sealed off with the pressure again at 5×10^{-8} mm. The tubes were then placed on life test for various periods of time and at various cathode temperatures. No emission current was drawn from any of the experimental cathodes included in this report.

After aging for a definite period of time the tubes were broken open, the oxide coatings were removed mechanically and x-ray diffraction patterns were made of the exposed interface. In order to be sure the interface compound was stable in air, the first samples were opened in a dry-box and then coated with a protective coating of waterproof wax. No change in the interface pattern was noticed between the cathodes treated in this manner and those which were not. The x-ray powder diffraction patterns were made with a camera of 7.16 cm radius and using Cu $K\alpha$-radiation filtered through 0.0007 in. nickel.

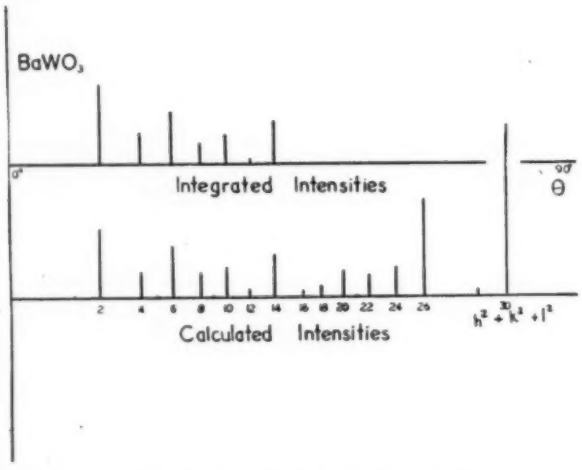


FIG. 1. The integrated intensities and Bragg angles of the x-ray powder diffraction lines of $BaWO_3$ compared with their calculated values.

The x-ray diffraction lines showed the interface compound to have a cubic structure with a lattice spacing of 4.30 kX. In addition to these and the tungsten lines, a few other lines were sometimes present which were probably due to tungsten oxides but they were always too weak to permit identification. Assuming the interface compound to have a perovskite ($CaTiO_3$) structure with a formula $BaWO_3$, the intensities were calculated and are shown in Fig. 1 together with the integrated intensities obtained by microphotometering the x-ray diffraction films. The calculation of these intensities involved the use of absorption factors for coated cylinders.[2] The lines beyond a Bragg angle of 45° were too broad to be measured. This broadening was probably due to the small crystal sizes and also to the fact that most perovskite type crystals are pseudo-cubic, causing the lines in the back reflections to have a multiple structure.

Cathodes consisting of SrO on W were also prepared and the x-ray analysis indicated the presence of $SrWO_3$ with a lattice constant of about 4.12 kX.** These lines were much weaker than those obtained from the barium compound, indicating the BaO reacts more easily with the W than does the SrO. Also prepared were cathodes with a coating of equal molar solid solution (BaSr)O on W. The lattice constant of the resultant $(BaSr)WO_3$ was 4.24 kX, indicating greater than an equal molar fraction of barium. Similar results to the above have been obtained when the cathode was formed on a 3 mm diameter nickel sleeve containing 4.7 percent tungsten impurity. Commercial hearing aid tubes[3] using a 0.5 mil tungsten wire base for the oxide cathode have been examined and the tungstite interface was observed to be present both in new tubes and in tubes which had been operated for 500 hours.

In order to determine whether the reaction requires the presence of any other material than the BaO and W, a special tube was constructed in which the BaO was deposited onto clean W by evaporation. $BaCO_3$ was sprayed onto a small spiral of pure platinum wire[4] which was then mounted in the vacuum tube with the tungsten wire passing along the axis of the spiral. First the $BaCO_3$ was reduced to BaO by passing current through the Pt wire. The W wire was then cleaned by raising its temperature to 1800°C. BaO was evaporated onto the cold W by raising the temperature of the Pt to about 1400°C until no oxide remained visible on the Pt. The W wire with its coating of BaO was heated to 850°C for 24 hours after which the cathode was removed and placed in the x-ray diffraction camera. X-ray diffraction patterns showed that $BaWO_3$ had been formed as in the previous tubes.

Several attempts were made to prepare $BaWO_3$ in larger quantities by heating mixtures of various tungsten and barium compounds in a vacuum furnace as for example powdered tungsten and barium carbonate. All these experiments resulted in forming the tungstate $BaWO_4$ instead of the tungstite $BaWO_3$. The lowest pressures obtainable in the vacuum furnace were of the order of 10^{-5} mm of Hg. When a mixture of powdered W and $BaCO_3$ was coated on a cathode sleeve and heated in a vacuum of about 10^{-7} mm the x-ray analysis showed a large percentage of $BaWO_3$ present and no $BaWO_4$. From these experiments it seems probable that the tungstite is formed only in a good vacuum and at higher pressures the tungstate is formed.

An investigation of chemical properties of $BaWO_3$ has not yet been carried out. However, it has been observed that although the tungstite is relatively stable when exposed to air over a period of several months, if a sample is placed in water for several minutes much of the tungstite oxidizes over into the tungstate.

A series of identical tubes were prepared and life tested at a temperature of 875°C in order to determine the variation in interface thickness with life. An x-ray method was employed for these thickness measurements.[2] The times recorded include the

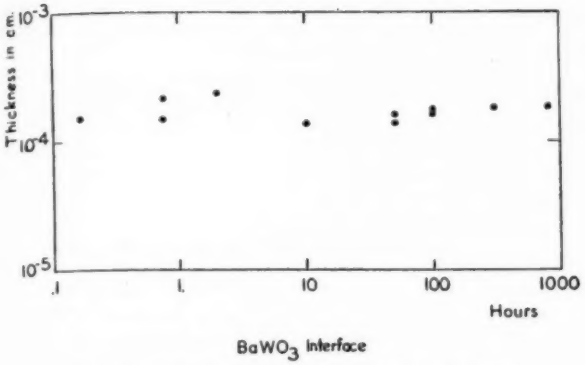


FIG. 2. Variation of interface thickness as a function of time at a temperature of 875°C.

time required to reduce the $BaCO_3$ to the oxide. Results are presented in Fig. 2 for the 15 cathodes examined. It can be seen that at this temperature the interface reaches its final thickness during the conversion process (less than 10 minutes) and remains essentially constant at 1.5×10^{-4} cm during the subsequent life of the tube. It should be pointed out that if the tungsten is present only as an impurity in another metal such as nickel, the rate of diffusion of the W may lead to an increase in thickness with life. Also drawing emission current from the cathode may well influence the interface thickness.

We wish to extend our thanks to Professor A. S. Eisenstein for many helpful discussions, and also to Mr. Harold John, who carried out the chemical preparations and to Mrs. Hope Lorch, who constructed most of the tubes.

* This work supported in part by ONR.
[1] A. S. Eisenstein, *Advances in Electronics I* (Academic Press, Inc., New York, 1948), p. 24.
[2] E. B. Hensley, Phys. Rev. **77**, 744(A) (1950).
** The *kX* unit used is about the same as 10^{-8} cm.
[3] Type CK522AX supplied by I. E. Levy of Raytheon Manufacturing Company.
[4] Prepared by H. T. Reeve of Bell Telephone Laboratories, Inc.

An Alternative Method for the Summation of Fourier Series

I. F. MORRISON
Department of Civil and Municipal Engineering, University of Alberta, Edmonton, Alberta, Canada
May 25, 1950

AS an alternative to the interesting method for the summation of certain types of Fourier Series presented in the April issue by L. A. Pipes, a more direct approach can be made by using summation formulas for trigonometric series in terms of the Bernoulli and Euler polynomials. There are four of these formulas, which are, perhaps, not quite as well known as they might be, for they are generally quite useful, especially in such cases as are referred to in this paper.

One of them, for example, is,

$$\bar{B}_{2\nu+1}(x)=(-1)^{\nu+1}\frac{2(2\nu+1)!}{(2\pi)^{2\nu+1}}\sum_{n=1}^{\infty}\frac{\sin 2\pi nx}{n^{2\nu+1}}$$

in which $\bar{B}_{2\nu+1}(x)$ is the Bernoulli polynomial in the interval $0\leqq x<1$. The series is absolutely and uniformly convergent for values of $\nu>0$, but not for $\nu=0$ in the neighborhood of the points $x=\pm0, \pm1, \pm2, \cdots$. To illustrate, consider Example II of the paper. If one puts $\nu=0$ in the expression above, there results

$$-\bar{B}_1(x)=+\frac{1}{\pi}\sum_{n=1}^{\infty}\frac{\sin 2\pi nx}{n}$$

which is in the same form as

$$F_2(t)=\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\left(\frac{n\pi t}{s}\right),$$

Eq. (4.7) of the paper with $2x=t/s$. $-\bar{B}_1(x)=\frac{1}{2}-x$ is a periodic function with period 1. The graph starts with a value $+\frac{1}{2}$ at $x=0$ and becomes discontinuous at $x=+1$ where it experiences a discontinuity $+1$, after which the cycle is repeated. Thus, the graph, with ordinates multiplied by 2, is the same as that of Fig. 2 of the paper.

The other examples in the paper can be dealt with in the same way.

Creep-Time Law for Zinc Crystals

E. P. T. TYNDALL
Department of Physics, State University of Iowa, Iowa City, Iowa
May 25, 1950

THE creep of suitably oriented zinc single crystals has been found to follow a very simple empirical law: $S=At^M$, in which S is the strain (non-elastic) occurring in time t, counted as zero at the instant of applying the load, and A and M are constants. The crystals have been tested under constant load and therefore constant stress since the reduction in area is negligible. Series of runs have been made in which the crystal was first loaded to just beyond the linear (elastic, or Hooke's law) region and then allowed to creep. The load was then increased by a small increment and a second run made; a further increase in load produced a third run and so on for five or six runs. The constant M, which is the slope of the straight line obtained for each run when plotted as log strain against log time, is about 0.5, varying somewhat from this mean value in different runs. The values of the constant A in any one series of runs lie on straight lines when plotted as log$A$ against the stress. The positions of the lines on such a plot depend on the history of the specimen and the purity of the zinc. Their slopes are nearly the same. The simple law has been found only with complete consistency for crystals in which one of the translation directions lies very close (within about 2°) to the direction of the projection of the length of the specimen into the basal (slip) plane, or, otherwise stated, the angle between slip plane and length of specimen is the same as the angle between the length and the nearest translation direction. The law appears to hold quite accurately for series of runs extending from strains of 10^{-7}, about the least observable strain, to strains somewhat above 10^{-3}. There is no indication of a separation into the two processes generally heretofore recognized, transient and steady-state creep.

For crystals with the translation direction 5° or more out of line with the projection of the length the behavior was definitely different. Some followed the pattern of transient plus slow creep found previously[1] while others had occasional large jumps, or (in one case) stepwise creep in which rapid creep followed slow creep at regular intervals. Crystals were also found in which, after a time of rapid creep, further strain followed the creep law above but with the constant M quite small, about 0.1 or 0.2. This appears to be typical of crystals with the angle mentioned above between approximately 15° and 30°. The simple creep law may be considered as a special case of the Andrade law proposed some years ago for creep in polycrystalline lead and copper.

[1] T. A. Read and E. P. T. Tyndall, J. App. Phys. **17**, 713 (1946).

Fracture Modes in High Purity Metals

T. J. AGNOR AND M. E. SHANK
Massachusetts Institute of Technology, Cambridge, Massachusetts
May 29, 1950

IT has been commonly accepted in many quarters that only iron and ferritic steels undergo a change in impact fracture mode with decreasing temperature. In such materials fracture changes from ductile type to brittle cleavage type and the absorbed energy drops markedly. The exact temperature of such

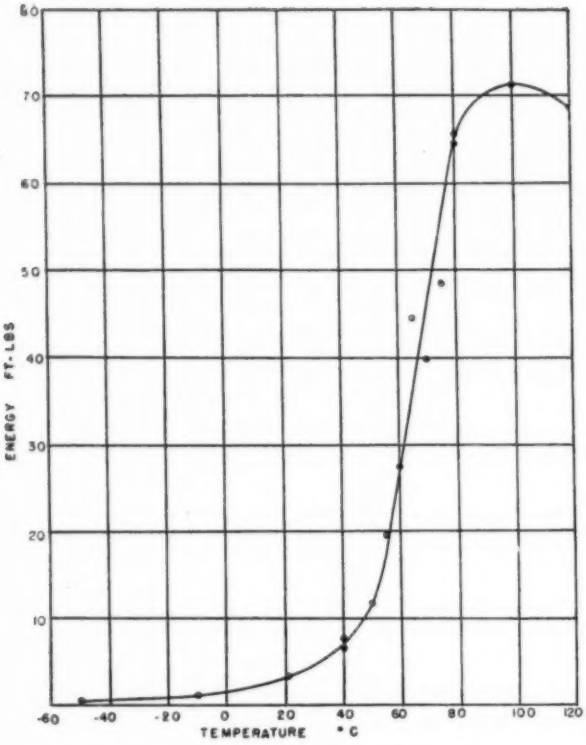


FIG. 1. Transition curve for high purity polycrystalline zinc.

transition depends upon composition and heat treatment of the material. Investigation by the authors into the behavior of high purity zinc by means of standard Charpy tests shows similar behavior of this metal. The material used was of 99.9955 percent purity, containing 0.0021 percent Pb, 0.0017 percent Cd, 0.0006 percent Fe, and 0.0001 percent Sn. It was hot-worked to an average grain diameter of 0.013 mm. The transition curve is shown in Fig. 1. The fractured surfaces, Fig. 2, have a macroscopic appearance that is superficially indistinguishable from that found in ferritic steels. Moreover, microscopic examination of the zinc fractures indicates brittle cleavage facets adjacent to large areas of ductile fracture in specimens whose gross macroscopic fracture appearance is entirely ductile. In some instances, large, deep voids have been opened in the material, apparently due to the triaxial stress conditions developed during fracture. It is evident, therefore, that transition temperatures exist for high purity metals other than iron and ferritic steels, and in fact for a metal not even of the body-centered lattice type.

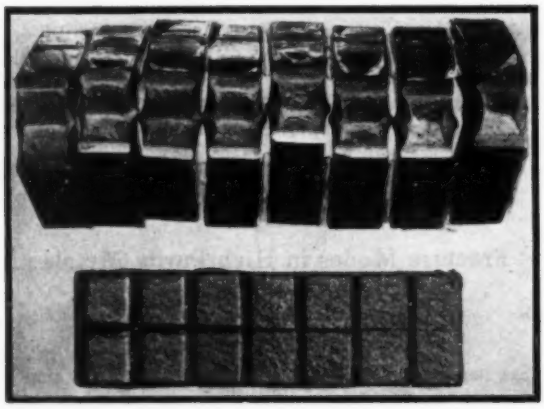


FIG. 2. Photograph of specimens showing fractured surfaces. Order of specimens, from highest to lowest temperature, is right to left on upper row, then left to right on lower row.

Similar work was carried out on high purity, vacuum-cast iron, major impurities in which were 0.063 percent O, 0.00038 percent H, 0.002 percent C. Nitrogen was not detectable. The iron had an average grain size of 0.31 mm, grain refinement being particularly difficult in such material. A very sharp transition temperature was found at 80°C. This is far above the usual transition ranges reported for ordinary Armco Iron,[1] and it does not seem probable that the difference is due entirely to the coarse grain size. On the basis of carbon and nitrogen present it would seem likely to expect a transition range in the neighborhood of −40° to −100°C. That it does not occur here apparently indicates the effect of the relatively high oxygen content. Since the work of Low and Gensamer[2] indicates that oxygen has no effect on aging or yield point in steels, the present evidence would serve to support the belief[3] that the mechanism of the transition temperature in metals is entirely unrelated to aging and inhomogeneous yielding phenomena.

[1] H. W. Gillett, "Impact Resistance and Tensile Properties of Metals at Subatmospheric Temperatures," ASTM Pamphlet, August 1941, p. 25.
[2] J. R. Low, Jr. and M. Gansamer, "Aging and the Yield Point in Steel," Trans. AIME **158**, 207 (1944).
[3] E. Orowan, "Fracture and Strength of Solids," Reports on Progress in Physics **XII**, 1949.

Intensity Correction Factors for X-Ray Spectrometer Transmission Pole Figure Determinations

SUZANNE VAN DIJKE BEATTY
Chemical Department, Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania
May 25, 1950

A TABLE has been computed to expedite the reduction of intensity contours obtained with a recording x-ray spectrometer using the transmission method for pole figure determination. This method is particularly suited for thin sheet material having very small grain size and is valuable in the determination of the textures of rolled materials.

In this type of pole figure work, the counter is held stationary at a given value of 2θ corresponding to a given reflection from the sample. The sample is then rotated about the vertical axis making an angle α with the Bragg angle. For each value of α, the sample is in turn rotated about the horizontal axis making an angle β with the rolling direction. Intensity corrections are necessary since the path length in the sample varies with different values of α and θ. The correction factor at any θ is taken as unity when $\alpha=0$. It is clear that no correction is needed for different values of β since the path length in the sample remains unchanged. The final intensity depends also on the absorption of the sample and on its thickness. Fortunately we need only the product μt so that the absorption factor can be determined experimentally from the relation $e^{-\mu t}=I/I_0$. The intensity of a strong reflection from quartz, for example, is recorded both with and without the sample to be studied in front of the counter slit. The ratio of the two intensities gives the required absorption factor. The correction factor also involves the diffracting angle θ since the path length and total volume of the sample irradiated differ.

The equation giving the intensity correction factor is given by:[1]

$$\frac{I_0}{I_{\pm\alpha}}=\frac{\mu t\exp(-\mu t/\cos\theta)}{\cos\theta}\times\frac{(\cos(\theta\pm\alpha)/\cos(\theta\mp\alpha))-1}{\exp(-\mu t/\cos(\theta\pm\alpha))-\exp(-\mu t/\cos(\theta\mp\alpha))}. \quad (1)$$

This is a special case of the expression given by Smoluchowski and Turner[2] and it was found[1] that more exact values are obtained when α is negative, i.e. with clockwise rotation of the shaft of the sample holder.

The factors are given for every 5° of α from 5° to 60° inclusive. For each value of α, every 10° in 2θ is listed between 10° and 80° and every 0.1 interval of the absorption factor from 0.1 to 0.9. Families of curves are given for every 5° of α in Figs. I to XII.

The tables may serve not only as a source of correction factors, but as a guide in the preparation of the sample and data. By plotting the factors as a family of curves for each value of θ or of $e^{-\mu t}$, one can see at a glance which set of values will require the least correction. The problem of optimum settings is also discussed by L. G. Schulz[3] in a recent paper.

A limited number of these tables is available for free distribution from the Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania. Requests for the tables should be addressed directly to the author.

[1] Decker, Asp, and Harker, J. App. Phys. **19**, 388–92 (1948).
[2] R. Smoluchowski and R. W. Turner, Rev. Sci. Inst. **20**, 173 (1949).
[3] L. G. Schulz, J. App. Phys. **20**, 1033–6 (1949).

Determination of Elastic Constants in Single Crystals with Especial Reference to Silver Chloride

DAVID L. ARENBERG
U. S. Naval Air Development Center, Johnsville, Pennsylvania
June 1, 1950

USING the ultrasonic pulse method[1] to measure the velocity of sound at 15 mc/sec. in different directions, the elastic constants of silver chloride have been determined. Two single crystals, F-16-a-2 and F-16-a-7, were oriented and ground by West and Makas of the Polaroid Corporation. After annealing carefully, the crystals were developed to produce a silver electrode on the surfaces.[2] Corrections were applied to account for the thickness of the piezoelectric transducers, but due to the large temperature and pressure coefficients of silver chloride, complete agreement between different sets was not obtained. No short time effect on the velocity was noticed when a strong light shone on the crystals. The sound absorption coefficient was low and hence not measured. The values for F-16-a-2 are believed to be better than for the set from F-16-a-7. The value 5.589 g/cc determined by Bridgman[3] from x-ray data was used for the density rather than the handbook value of 5.56.

Selected values for crystal F-16-a-2 are 0.605, 0.364, and 0.0624 times 10^{12} dynes/cm^2 for C_{11}, C_{12}, and C_{44} respectively.

Since it is not always possible to obtain orientations of crystals that allow simple calculations between the elastic constants and velocities, approximate formulas were developed and the measurements in the *hkl* directions made to check them. One can derive explicit formulas that are very good first-order approximations for all orientations and exact in certain planes and points, such as all directions in the plane normal to an 001 axis and also in the 111 direction for cubic crystals. One uses the direction cosines involving rotations about three axes with the IRE conventions given by Cady[4] of angles φ, θ, and ψ. φ and θ have their usual significance of azimuth and colatitude in spherical coordinates but ψ, which represents rotation about the Z''' axis, will be indeterminate for a pulse traveling in an infinite medium, and must be either eliminated or associated with some properties of the waves to have physical significance.

The choice of coordinates used by the experimenter is an arbitrary one which cannot influence the velocity in the medium. It is observed that if the piezoelectric transducers, particularly those generating complex waves, are mounted on the specimen the time delay between members of the series of echoes received from a single pulse will differ by increments which are integral multiples of a linear combination of the delays due to any of the three different modes having different velocities traversing the specimen once. Even with a pure mode generator operating at normal incidence, mode transfer can occur at each reflection resulting in a complex system of echoes which can, however, be analyzed into the three different fundamental modes. Changing the orientation of a shear wave transducer on the specimen can only influence the relative amplitudes of the echoes. With sufficient care in orientation, only one period between echoes will be observed.

One can therefore associate the angle ψ with the plane of polarization of the quasi-transverse waves in the crystal and by *ad hoc* reasoning apply Fermat's principle that the velocities of the waves assume stationary values so that all $\partial C_{ii}/\partial\psi=0$. The angle ψ is then uniquely determined as a function of φ and θ and will give the velocities and plane of polarization to a first-order approximation as shown below. Higher order correction terms can be added later.

One selects the expressions for C_{44}', C_{55}', and C_{33}' for operation and defining for cubic crystals:

$$A=8-\sin^2\theta(7+\cos^4\varphi) \tag{1}$$

$$B=2\cos\theta\sin^4\varphi \tag{2}$$

$$C=A-2(1-\cos^4\varphi) \tag{3}$$

$$\tan 2\psi=B/C \tag{4}$$

$$r=C_{11}-C_{12}-2C_{44}, \tag{5}$$

we can get the elastic constants along the Z''' axis:

$$C_{33}'=C_{11}-r\frac{A\sin^2\theta}{4}=C_{11}-r\alpha \tag{6}$$

$$C_{44}'=C_{44}+r\frac{\sin^2\theta}{8}(A-(B^2+C^2)^{\frac{1}{2}})=C_{44}+r\beta \tag{7}$$

$$C_{55}'=C_{44}+r\frac{\sin^2\theta}{8}(A+(B^2+C^2)^{\frac{1}{2}})=C_{44}+r\gamma. \tag{8}$$

Also for any direction in the crystal, the velocities are:

$$V_L'=(C_{33}'/\rho)^{\frac{1}{2}} \tag{9}$$

$$V_t'=(C_{44}'/\rho)^{\frac{1}{2}} \tag{10}$$

$$V_T'=(C_{55}'/\rho)^{\frac{1}{2}}. \tag{11}$$

After determining all C_{ii}' from measurements and manipulating (6), (7), and (8), one gets the inverse relations:

$$C_{11}=C_{33}'+\frac{\alpha(C_{55}'-C_{44}')}{\gamma-\beta} \tag{12}$$

$$C_{12}=C_{33}'-C_{44}'-C_{55}'-(1-2\alpha)\frac{(C_{44}'-C_{55}')}{\gamma-\beta} \tag{13}$$

$$C_{44}=\frac{\gamma C_{44}'-\beta C_{55}'}{\gamma-\beta}. \tag{14}$$

If φ and θ are known with respect to the crystal axes the three velocity measurements in one direction, such as the 110, would suffice to give all three constants except for a certain arbitrariness in the relation between γ, β, and C_{44}' and C_{55}'. For most ionic crystals, the anisotropy factor $(C_{11}-C_{12})/2C_{44}$ is >1.00 and for metallic crystals <1.00, so that this choice can be made. Errors in measurement and misorientation are least important in the 001 and 111 directions, but here $V_t'=V_T'$ and velocities must be measured in two directions.

Furthermore, with highly anisotropic crystals, one can obtain an approximate idea of the orientation of the crystal by measuring the transverse velocities in three or more directions. The

TABLE I. Orientation and path lengths.

	001	110	*hkl*	001	110	*hkl**
φ	~	44½°	126½°	135°	45°	124°
θ	0°	89½°	97½°	1°	90°	83½°
Length	1.092″	0.755″	.936″	.943″	.630″	.967″
	Crystal F-16-a-2			Crystal F-16-a-7		

* The *hkl* surface is near a plane of the form 531 which in turn is near the center of the unit triangle 100-110-111.

TABLE II. Velocity measurements $\times 10^{-6}$ cm/sec.

Mode \ Face	001	110	*hkl*	001	110	*hkl*
V_L'	.3294	.3125	.3135	.3276	.3116	.3119
V_T'	.1056	.1465	.1423	.1055	.1455	.1406
$V_L'V_T'$	.1057	.1056	.1083	.1055	.1054	.1080
	Crystal F-16-a-2			Crystal F-16-a-7		

TABLE III. Elastic constants of silver chloride ($\times 10^{-12}$ dynes/cm^2).

Const. \ Face	001	110	*hkl*	001	110	*hkl*	Ave.
C_{11}	.606	.603	.603	.600	.599	.593	.601
C_{12}	—	.364	.365	—	.362	.356	.362
C_{44}	.0624	.0623	.0627	.0622	.0621	.0631	.0625
$(C_{11}'+C_{44}'+C_{44}')$	.731	.728	.728	.724	.723	.720	.726
$\frac{(C_{11}+2C_{12})^*}{3}$	—	.444	.444	—	.441	.436	.441
Dimensionless isotropy ratio	—	1.92	1.90	—	1.90	1.88	1.90
	Crystal F-16-a-2			Crystal F-16-a-7			

* Previously reported values for the bulk modulus are 0.417 reported by Richard and Jones, J. Am. Chem. Soc. **31**, 158 (1909) and a value of 0.412 obtainable by extrapolating Bridgman's data to zero pressure.

ratio of the difference of the pairs of transverse waves in directions (a) and (b) is a function of the angles φ and θ alone as:

$$\frac{(V_t'^2-V_T'^2)_a}{(V_t'^2-V_T'^2)_b}=\frac{((B^2+C^2)^{\frac{1}{2}}\sin^2\theta)_a}{((B^2+C^2)^{\frac{1}{2}}\sin^2\theta)_b}. \tag{15}$$

Contours of the $F(\varphi,\theta)=\log(\sin^2\theta(B^2+C^2)^{\frac{1}{2}})$ in stereographic projection have been made to facilitate such operations.

[1] D. L. Arenberg, "Supersonic Solid Delay Lines," M.I.T. Rad. Lab. Report No. 932, February 1946.
[2] J. R. Haynes, Rev. Sci. Inst. **19**, 51 (1948). (Note: Use 10 times the total amount of water.)
[3] P. W. Bridgman, Proc. Am. Acad. Arts Sci. **74**, 21–25
[4] W. G. Cady, *Piezoelectricity* (McGraw-Hill Book Company, Inc., New York, 1948), p. 83.

Examination of Thin Overgrowths by Multiple Scattering of Electrons

L. G. SCHULZ
Institute for the Study of Metals, University of Chicago, Chicago, Illinois

June 14, 1950

ELECTRON diffraction by the reflection method is commonly employed in the study of oriented overgrowth on single crystals.[1] In Fig. 1A is shown the usual experimental arrangement. When the overgrowth is produced by evaporation and deposition in a vacuum the deposit forming the overgrowth is occasionally too small to be detected easily. To enhance the scattering of very small deposits the angle θ might be decreased. This leads to difficulties, however, because of refraction effects and increased background scattering. It has been found that a better procedure is to employ the reflected beam from the substrate as the incident beam on the deposit. In drawing B of Fig. 1 the part of the incident beam I which is reflected is indicated by I'. Since I' is an incident beam in respect to the deposit there will be a second diffraction process resulting in the beams I''. The angle θ' is associated with the crystal constants of the deposit.

With the arrangement of drawing B one can take advantage of much smaller scattering angles for the deposit than by the direct approach of drawing A. The part of the beam I (in drawing B) which is intercepted by the deposit must be scattered through a total angle of $2\theta\pm\theta'$ to reach the same general area of the photograph as the beams I''. Because of the rapid decrease in intensity with angle of scattering only the spots due to I'' (and to I''') appear in the photograph.

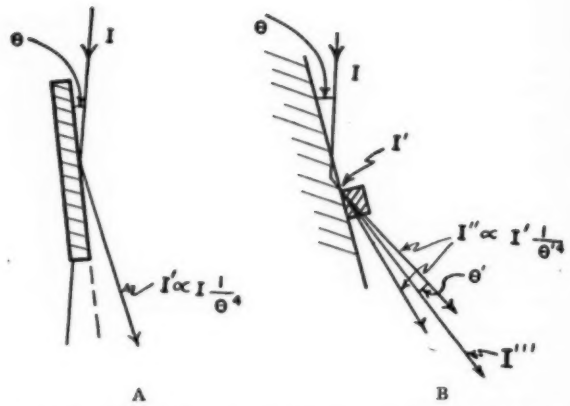
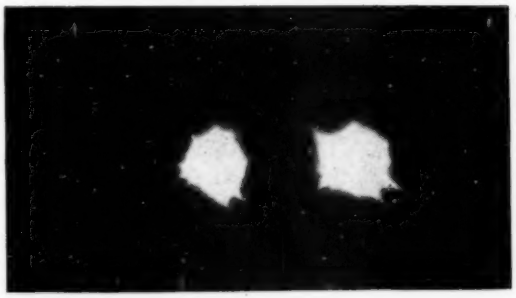


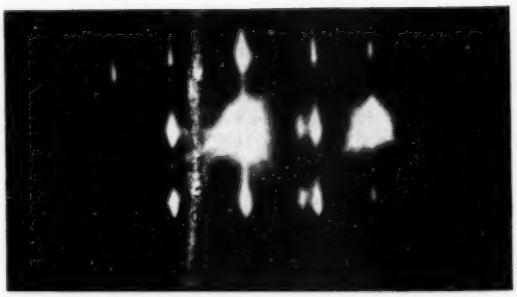
FIG. 1. Experimental arrangements. Drawing A shows the usual reflection arrangement; drawing B shows the new method employing multiple scattering of electrons.

The results from an application to a study in polymorphism are presented in Fig. 2. Photograph (a) is from a single crystal of LiF, the two points being the (600) and (620) reflections. Photograph (b) is from the same crystal but after a deposit of CsBr averaging 2A in thickness has been added. This small deposit has migrated over the substrate and grown into crystals approximately 100A on an edge. An analysis of photograph (b) reveals that the CsBr crystals have a NaCl type structure, rather than the usual CsCl type, and that they are oriented with their axes parallel to the corresponding axes of the substrate.

In Fig. 2 each of the two beams from the LiF substrate is an incident beam in respect to the deposit. As a result there are two secondary diffraction patterns. The displacement of the two patterns relative to each other is the same as the separation of the two LiF points. Consequently the separation of any pair of corresponding points from the two secondary patterns can be used as



(a)



(b)

FIG. 2. Results of an app' cation of the new method. CsBr has been deposited by evaporation and deposition in a vacuum onto a single crystal of LiF. (a): LiF (heavy exposure); (b): 2A of CsBr added to LiF.

an internal standard making possible an accurate determination of the crystal constants of the deposit.

Details concerning the range of application, together with additional results, will be given later.

This research was supported in part by Army Air Force Contract No. AF 33 (638)-6534.

[1] G. P. Thomson and W. Cochrane, *Theory and Practice of Electron Diffraction* (Macmillan and Company, London, 1939), Chapter XI.

A Note on Thermoelectric Generation of Current

MAX B. GREEFF
Upper Montclair, New Jersey
July 5, 1950

TELKES wrote a very thorough and interesting article[1] about the efficiency of a thermoelectric generator. A formula is given for the calculation of the efficiency for any thermocouple whenever the Wiedemann-Franz-Lorenz law is applicable. The calculations given in Table I of the article show that the efficiency of a thermocouple may become much better than has been anticipated in the past, provided we have a couple delivering power beyond 200 microvolts per °C. In fact, at 500 microvolts per °C we may expect an efficiency of 20.9 percent if the temperature gradient between hot and cold junctions is 500°C.

However, there is a limitation in Telkes' formula to which Papet[2] calls our attention. The electrical efficiency, depending on the relation of internal and external resistances, is fixed at 50 percent (internal and external resistances are assumed equal); this being the condition of maximum energy output of the thermoelectric generator. Though we are much concerned in obtaining high energy, it is not always most efficient to operate the generator at maximum energy output.

For example, let us calculate a complete characteristic of a generator assuming 4 different cases of construction: (1) applying thermocouple elements of 100 microvolts per °C; (2) applying thermocouple elements of 500 microvolts per °C; (3) changing the relation of internal and external resistances of case (2); (4) reducing the number of elements in case (3).

The energy output of all 4 cases remains the same (Table I).

TABLE I.

Generator case	*I*	*II*	*III*	*IV*
Power of each couple microvolts per °C	100	500	500	500
Generator e.m.f. volts	100	100	100	55.5
Number of couples required	2000	400	400	222
Voltage utilized volts	50	50	95	50
Amperage utilized amp.	25	25	13.2	25
Wattage utilized kilowatts	1.25	1.25	1.25	1.25
Resistance intern. ohms	2	2	0.4	0.22
Resistance extern. ohms	2	2	7.2	2.00
Resistance total ohms	4	4	7.6	2.22
Resistance per couple ohms	0.001	0.005	0.001	0.001
A efficiency Telkes %	2.34	20.9	—	—
B electrical $\frac{r_2}{r_1+r_2}$ %	50	50	95	90
C eff. total $=2A\times B$ %	2.34	20.9	39.8	37.6

We see immediately that there is more than one way to construct a thermoelectric generator of a given output. At maximum output, Cases *I* and *II*, we may adopt the calculated efficiencies offered by Telkes. But Cases *III* and *IV* require discussion as they are not at maximum output. Consequently, we are not allowed to apply Telkes' efficiency for these cases. We may, however, compare the two cases *III* and *IV* with Case *I* which has the same resistance per element and only differs in the number of elements.

Case *III* has 1/5 the number of elements as has Case *I*, producing the same energy. Consequently, Case *III* has 5 times the efficiency as Case *I*, using only 1/5 the amount of fuel. The efficiency therefore is $5\times 2.34=11.60$ percent.

Case *IV* has 1/9 the number of elements as has Case *I*, producing the same energy. Consequently, Case *IV* has 9 times the efficiency as Case *I*, using only 1/9 the fuel. Therefore, the efficiency is $9\times 2.34=21.06$ percent.

We could construct a generator of the same elements and equal output, but at different voltage, whose efficiency is still higher, choosing the special Case *V* where the numerical figures for voltage and amperage are practically equal.

This Case *V* would require only 160 elements or thermocouples to deliver a voltage of 36 volts at an amperage of 35 amp. Using only 1/12.5 of the number of elements as in Case *I*, the efficiency is $12.5\times 2.34=29$ percent. Under these circumstances the generated e.m.f. would be 40 volts and the electrical efficiency 90 percent.

[1] M. Telkes, J. App. Phys. **18**, 1116 (1947).
[2] R. M. Papet, J. App. Phys. **19**, 1180 (1948).

On the Negligible Changes Produced in the Stress-Strain Curve by Immersion of a Copper Single Crystal in Mercury

J. B. WACHTMAN, JR.
Carnegie Institute of Technology, Pittsburgh, Pennsylvania
July 6, 1950

IN his proposed explanation of slip bands, Frank[1] considers internal processes and reflections of fast dislocations at the surface of a single crystal as dislocations of opposite sign. He suggests that immersing a single crystal in a liquid of higher density will suppress reflections of dislocations. If reflection of dislocations plays an appreciable role as compared with internal processes in the formation of slip bands, one might expect the suppression of reflections to change the stress-strain curve.

Single crystals of 99.999 percent pure copper were grown to A.S.T.M. standard tensile specifications (2-in. gauge length, 0.25-in. diameter). The crystals were oriented and then annealed in vacuum for 20 hours at 600°C.

These samples were then subjected to tensile tests in a Baldwin Southwark Tate-Emery 60,000-pound testing machine; the extension being determined from the travel of the head of the testing machine as determined by a dial gauge. This method of extension measurement was checked on three crystals by measuring the separation of two marks on the crystal initially 1 inch apart with two cathetometers. The stress and strain were resolved by assuming that single slip took place until a position of symmetry with respect to two slip systems was reached after which it was assumed that equal slip on both systems took place. Three crystals were pulled partly in air and partly under mercury. Table I gives, for

TABLE I.

Crystal No.	Final load	Resolved shear strain at beginning of double slip	Resolved shear strain at which mercury was added
3	1022 lb.	53.8%	No Hg added
4	764	48.4%	9.5%
6	768	29.5%	No Hg added
10	1012	52.1%	No Hg added
11	731	56.4%	0.0%
13	823	64.0%	10.3%

each crystal, the resolved shear strain at which double slip begins, the resolved shear strain at which mercury was added, and the final load.

The resolved stress-strain curves for the crystals strained in air are shown in Fig. 1 as calculated from the cathetometers and the dial gauge extensometer. The two methods of calculation yield curves in good agreement with each other for crystals 3 and 10. The curve for No. 6 computed from the dial gauge has an unusual bump in it and does not agree well with the other curve for No. 6 at intermediate strains. The strain determined from the

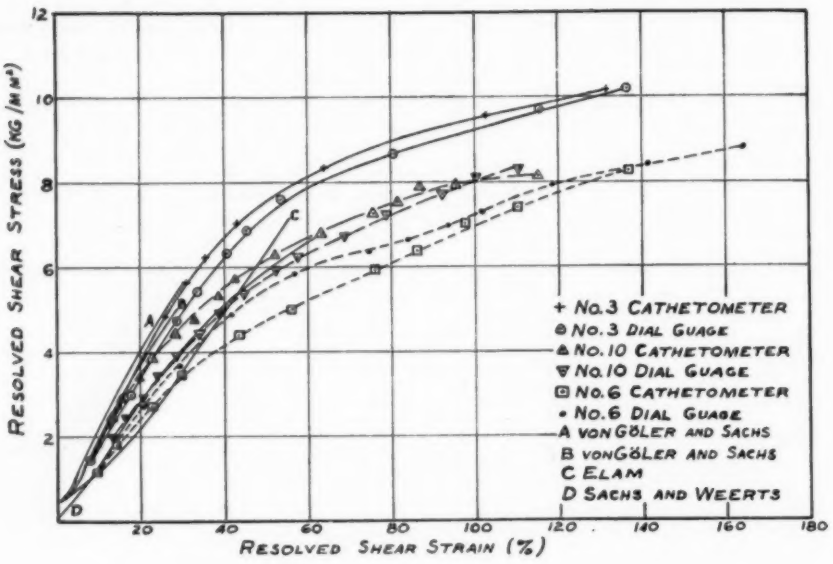


FIG. 1. Stress-strain curves obtained on copper crystals in air.

dial gauge may be in error because of inhomogeneous deformation near the grips, slipping of the grips in the machine, or sticking of the plunger in the dial gauge. Since the high strain data yields close agreement it seems reasonable to assume that the error at intermediate strain was caused by sticking of the dial gauge plunger and that the high and low strain portions of No. 6 curve, in agreement with No. 3 and No. 10, indicate that inhomogeneous deformation near the grips and slipping of the grips do not produce large errors in the strain computed from the dial gauge.

The curves for No. 4 and No. 13 lie well within the range of the curves run in air. The curve for No. 11 does drop below the lowest of the curves run in air but the greatest stress difference between it and the curve for No. 6 is 0.8 kg/mm² which is to be compared with a spread of 2.9 kg/mm² between the curves run in air. Thus if the weakness of No. 11 is ascribed to the mercury (which is improbable in view of the curves for No. 4 and No. 13) the effect is smaller than the normal variation in annealed, high purity crystals.

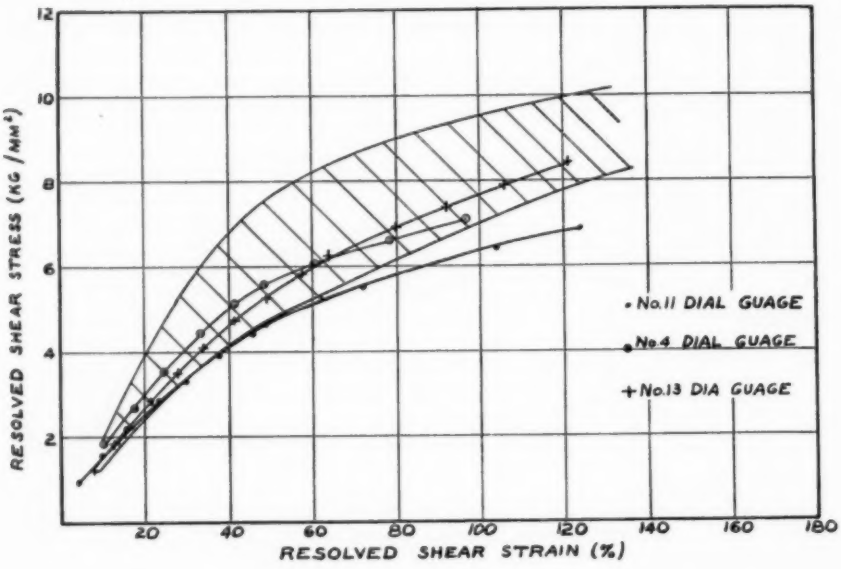


FIG. 2. Comparison of stress-strain curves obtained in air and in mercury. The shaded region gives the spread of the curves obtained in air.

Curves by von Goler and Sachs,[2] by Sachs and Weerts,[3] and by Elam[4] are also shown in Fig. 1.

The curves obtained from the mercury experiments are shown in Fig. 2. The shaded region indicates the spread of the curves obtained on specimens run in air.

The curvature and the roughness of the surface of the crystals made a microscopic examination of the slip lines impossible but to the naked eye the appearance of the slip lines was the same on crystals strained under mercury and those strained in air.

In conclusion it appears that immersion of copper single crystals in mercury does not have any important influence on the resolved stress-strain curve.

The author wishes to express his gratitude to Dr. J. S. Koehler for his aid.

[1] F. C. Frank, *Report of a Conference on Strength of Solids.*
[2] F. von Goler and G. Sachs, Zeits. f. Physik **55**, 581 (1929).
[3] G. Sachs and J. Weerts, Zeits. f. Physik **62**, 473 (1930).
[4] C. F. Elam, Proc. Roy. Soc. **A112**, 289 (1926).